

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 15:28:07 ; Search time 28 Seconds

(without alignments)
1482.779 Million cell updates/sec

Title: US-10-046-433-40

Perfect score: 1001

Sequence: 1 MAEPGSHHLSARVGRTER.....LGRSHNLPRGLMLDTQCR 1001

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 11892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 11892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 150 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	8	0.8	284 1	KIME_ARCFU
2	8	0.8	353 1	PERA_ARMFU
3	8	0.8	354 1	PERC_ARATH
4	8	0.8	443 1	TOLE_RICPR
5	8	0.8	4644 1	DYHC_MOUSE
6	7	0.7	73 1	NER_BPDI0
7	7	0.7	75 1	LSM6_SCHPO
8	7	0.7	80 1	LSM6_HUMAN
9	7	0.7	100 1	URE3_ECOLI
10	7	0.7	100 1	URE3_PROMI
11	7	0.7	100 1	URE3_PROVU
12	7	0.7	175 1	DEF_RICCN
13	7	0.7	175 1	Y773_METUA
14	7	0.7	181 1	YCF4_ODOSI
15	7	0.7	215 1	PMG2_ECOLI
16	7	0.7	239 1	DHSB_RECAM
17	7	0.7	244 1	IF6_ECHPO
18	7	0.7	267 1	TRY7_ANOGA
19	7	0.7	292 1	YCI5_PYRHO
20	7	0.7	293 1	RRP4_HUMAN
21	7	0.7	300 1	DPPC_ECOLI
22	7	0.7	320 1	STBA_ECOLI
23	7	0.7	346 1	YAD0_SCHPO
24	7	0.7	350 1	AROG_ECOLI
25	7	0.7	362 1	AROG_HAETN
26	7	0.7	439 1	AKR_ARATH
27	7	0.7	450 1	CBPA_SYNP7
28	7	0.7	451 1	TBA2_NEUCR
29	7	0.7	457 1	RADA_LISMO
30	7	0.7	482 1	CAB1_METUA
31	7	0.7	515 1	PEN3_ADESI
32	7	0.7	538 1	RO60_HUMAN
33	7	0.7	538 1	RO60_MOUSE

34	7	0.7	591 1	GV7_XENLA
35	7	0.7	603 1	MTRR_SCHPO
36	7	0.7	618 1	CAB2_METUA
37	7	0.7	624 1	NKX3_RAT
38	7	0.7	644 1	NKX3_HUMAN
39	7	0.7	645 1	NKX3_MOUSE
40	7	0.7	677 1	CEPA_TREPA
41	7	0.7	717 1	NH4_RAT
42	7	0.7	732 1	PRIA_ECOLI
43	7	0.7	740 1	EF2_PYRAE
44	7	0.7	777 1	UNC8_CABEL
45	7	0.7	892 1	RA16_SCHPO
46	7	0.7	944 1	YMH6_YEAST
47	7	0.7	1024 1	POP1_HUMAN
48	7	0.7	1040 1	MAN1_RAT
49	7	0.7	1254 1	POLS_EEVR8
50	7	0.7	1254 1	POLS_EEVRM
51	7	0.7	1254 1	POLS_EEVRT
52	7	0.7	1255 1	POLS_BEVY3
53	7	0.7	1255 1	POLS_BEVYP
54	7	0.7	2514 1	POLN_SINDO
55	7	0.7	2549 1	FRAP_HUMAN
56	7	0.7	2554 1	TLES_DROME
57	7	0.7	2594 1	TLES_DROVI
58	7	0.7	2845 1	APC_MOUSE
59	7	0.7	3829 1	SACS_HUMAN
60	7	0.7	3830 1	SACS_MOUSE
61	7	0.7	4644 1	DYHC_RAT
62	7	0.6	37 1	PSBL_ARATH
63	7	0.6	37 1	TX21_SELHU
64	7	0.6	37 1	TX22_SELHU
65	7	0.6	38 1	PSBL_CHIVU
66	7	0.6	38 1	PSBL_MARPO
67	7	0.6	43 1	P1V6_ADEB2
68	7	0.6	48 1	ATP8_YARLI
69	7	0.6	54 1	ATP8_PARLI
70	7	0.6	67 1	Y737_ARCFU
71	7	0.6	78 1	RUXF_SCHPO
72	7	0.6	79 1	NSGX_HUMAN
73	7	0.6	81 1	YVKA_VACCC
74	7	0.6	83 1	Y67_BPT3
75	7	0.6	86 1	Y425_TREPA
76	7	0.6	88 1	FXY4_MOUSE
77	7	0.6	89 1	XHLA_BACSU
78	7	0.6	89 1	YALI_BACLI
79	7	0.6	95 1	ES6B_MYCTU
80	7	0.6	97 1	XPA_CRIGR
81	7	0.6	97 1	YAF1_BACLI
82	7	0.6	98 1	RK23_ASTLO
83	7	0.6	102 1	VNI6_YEAST
84	7	0.6	105 1	GNM1_METTM
85	7	0.6	105 1	HIS3_LISIN
86	7	0.6	105 1	THI1_SYNY3
87	7	0.6	108 1	Y793_SYNY3
88	7	0.6	109 1	VIF_HVISC
89	7	0.6	112 1	RLAI_DROME
90	7	0.6	113 1	V195_ROMPY
91	7	0.6	115 1	HIS3_CLOAB
92	7	0.6	120 1	VP3_CAV26
93	7	0.6	121 1	SR14_ARATH
94	7	0.6	121 1	VP3_CAV82
95	7	0.6	121 1	VP3_CAVCI
96	7	0.6	121 1	VP3_CAVCI
97	7	0.6	124 1	VCI1_PEA
98	7	0.6	127 1	ACPS_THETN
99	7	0.6	128 1	CD59_AOTTR
100	7	0.6	128 1	V105_VACCV
101	7	0.6	128 1	V105_VARY
102	7	0.6	130 1	CLP1_DROME
103	7	0.6	130 1	VHEM_MNV
104	7	0.6	131 1	IMEP_STRNI
105	7	0.6	133 1	ACPS_CLOPE
106	7	0.6	133 1	RLI5_HELPJ

P20398	xenopus lae
Q10258	schizosacch
Q58776	methanococ
Q96901	rattus norv
Q96C58	homo sapien
Q99P47	mus musculu
O56336	treponema p
P26434	rattus norv
P17888	eschierichia
Q822C1	pyrobacillu
Q19174	caenorhabdi
P36617	schizosacch
O03631	saccharomyc
O99575	homo sapien
P21139	rattus norv
P05674	venezuelan
P36331	venezuelan
P09592	venezuelan
P36332	venezuelan
P27283	sindbis vir
P42345	homo sapien
P13368	drosophila
P20806	drosophila
O61335	mus musculu
O92454	homo sapien
Q91168	mus musculu
P38650	rattus norv
P29301	arabidopsis
P82959	selenococci
P82960	selenococci
P63139	chlorella v
P12165	marichantia
Q96627	bovine aden
Q36257	yarrowia li
P12697	paracentrot
O29521	archaeoglob
O59734	schizosacch
Q94864	homo sapien
P20569	vacchina vi
P20330	bacteriopho
O34430	treponema p
O92420	mus musculu
P39798	bacillus su
O99164	bacillus li
O05454	mycobacteri
O64029	cricetulus
P37135	bacillus li
P34771	astasia lon
P48232	saccharomyc
O58786	methanobact
Q92689	listeria in
P52232	synecocyst
O55933	synecocyst
P05899	human immun
P08570	drosophila
O91538	fowlpox vir
Q91KH7	clostridium
P54095	chicken ane
O04421	arabidopsis
P54096	chicken ane
O99152	chicken ane
P54094	chicken ane
P02856	pisum sativ
Q84857	thermoanaer
P51447	actus trivi
P07615	vacchina vi
P33043	variola vir
P02819	narcissus m
P15097	streptomyce
P01077	streptomyce
Q8XNPI	clostridium
O9XJ58	helicobacte

107 6 0.6 134 1 VRIM_ECOLI
108 6 0.6 135 1 RL15_HELBY
109 6 0.6 135 1 SEPH_AQUE
110 6 0.6 139 1 ACPS_BRAVA
111 6 0.6 140 1 GTH2_ANGAN
112 6 0.6 141 1 LYSB_BPP2
113 6 0.6 142 1 VR72_SCHPO
114 6 0.6 142 1 YNK5_YEAST
115 6 0.6 143 1 ANG3_MOUSE
116 6 0.6 146 1 HBE_DIDMA
117 6 0.6 148 1 YMS3_CAEEL
118 6 0.6 148 1 YMS4_CAEEL
119 6 0.6 150 1 ARG2_CLOBE
120 6 0.6 150 1 VRHP_HAEIN
121 6 0.6 152 1 SPIA_STRPU
122 6 0.6 154 1 PYDA_AERPE
123 6 0.6 157 1 VR6_HPV36
124 6 0.6 159 1 RL1G_ARATH
125 6 0.6 160 1 YGE7_PSEAE
126 6 0.6 162 1 G160_BAGSU
127 6 0.6 162 1 STEB_ECOLI
128 6 0.6 163 1 DNE1_CHLRE
129 6 0.6 164 1 YOR2_AZOVI
130 6 0.6 165 1 TPX_HAEIN
131 6 0.6 166 1 PSAL_SYNP7
132 6 0.6 170 1 RL17_MYCLE
133 6 0.6 171 1 IPPI_HUMAN
134 6 0.6 173 1 AROK_BUCAI
135 6 0.6 175 1 IPVR_SALTY
136 6 0.6 176 1 IPVR_VIBCH
137 6 0.6 177 1 AAC1_PSEAE
138 6 0.6 177 1 ATPD_BUCAP
139 6 0.6 181 1 PANC_ECOLI
140 6 0.6 183 1 KITH_FOWPV
141 6 0.6 187 1 YKT9_YEAST
142 6 0.6 189 1 OXYR_MYCKE
143 6 0.6 190 1 SOMA_PAROL
144 6 0.6 195 1 PAPH_ECOLI
145 6 0.6 195 1 PRSH_ECOLI
146 6 0.6 196 1 HIR5_HUMAN
147 6 0.6 197 1 CCP1_CRIGR
148 6 0.6 197 1 MOBA_METTH
149 6 0.6 198 1 ATPF_MYCGA
150 6 0.6 199 1 HIR5_MOUSE

ALIGNMENTS

RESULT 1
KIME_ARCFU
ID KIME_ARCFU STANDARD; PRT; 284 AA.
AC 027995;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Mevalonate kinase (EC 2.7.1.36) (MK).
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
ON NCBI_TaxID=2234;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Richardson D.L., Dodson R.J., Gilm M., Hickey E.K., Peterson J.D.,
RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftis S.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,

P75781 escherichia
P56040 helicobacte
O66615 aquiflex aeo
O69159 bradyrhizob
P27767 anguilla an
P51770 bacteriopho
O14155 schizosach
P30943 saccharomyc
P97802 mus musculu
P11025 didelphis m
P34499 caenorhabdi
P34500 caenorhabdi
P56685 clostridium
P44675 haemophilus
O4109 strongyloce
P09428 aeropyrum p
P50810 human papil
O65569 arabidopsis
O98747 pseudomonas
P80872 bacillus su
P38392 escherichia
P57525 chlamydomon
O69219 azotobacter
O57549 haemophilus
P58822 synechococc
O84797 mycobacteri
Q13522 homo sapien
P57605 buchiera ap
O8X910 saimonella
O9K934 vibrio chol
P23181 pseudomonas
O51875 buchiera ap
P18103 escherichia
P10052 towplox vir
P60452 saccharomyc
O87883 mycobacteri
P09537 paralichthy
P07111 escherichia
P42185 escherichia
Q9um60 homo sapien
O35847 cricetus
O26246 methanobact
P33256 mycoplasma
Q9qz23 mus musculu

RA Cotton M.D., Spriggs T., Ariach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RT Nature 350:364-370(1997).
RL Nature 350:364-370(1997).
CC -1- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-
phosphomevalonate.
CC -1- COFACTOR: MAGNESIUM (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000946; AAB89365.1; -.
DR TIGR; AF2289; -.
DR InterPro; IPR001745; GHMPkinase_ATP.
DR InterPro; IPR001459; Mev_gal_kin.
DR Pfam; PF00288; GHMP_kinases; 1.
DR PRINTS; PR00959; MEVGALKINASE.
DR TIGRFAMs; TIGR00549; mevalon_kin; 1.
DR PROSITE; PS00627; GHMP_KINASES_ATP; 1.
KW Transferrase; Kinase; ATP-binding; Magnesium; Complete proteome.
FT NP BIND 86 ATP (POTENTIAL).
SQ SEQUENCE 284 AA; 30877 MW; F102C1C71772CA7F CRC64;
Query Match 0.8%; Score 8; DB 1; Length 284;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 593 LEASDVGS 600
Db 201 LEASDVGS 208
RESULT 2
PERA_ARMRU
ID PERA_ARMRU STANDARD; PRT; 353 AA.
AC P00433;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peroxidase C1A precursor (EC 1.11.1.7).
GN PRXC1A OR HPRC1.
OS Ammoracia rusticana (Horseneadish) (Ammoracia lappathifolia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid II; Brassicales; Brassicaceae; Ammoracia.
ON NCBI_TaxID=3704;
RX MEDLINE=88225087; PubMed=3371352;
RA Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,
RA Shimizu A., Takano M., Yamada Y., Okada H.;
RT "Structure of the horseadish peroxidase isozyme C genes.";
RL Eur. J. Biochem. 173:681-687(1988).
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88225087; PubMed=3371352;
RA Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,
RA Shimizu A., Takano M., Yamada Y., Okada H.;
RT "Structure of the horseadish peroxidase isozyme C genes.";
RL Eur. J. Biochem. 173:681-687(1988).
RN [2]
RP SEQUENCE OF 31-338.
RX MEDLINE=7068850; PubMed=1001465;
RA Welinder K.G.;
RT "Covalent structure of the glycoprotein horseadish peroxidase (EC
RT 1.11.1.7).";
RL FEBS Lett. 72:19-23(1976).
RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
 RX MEDLINE=98069652; PubMed=9406554;
 RA Gajhede M., Schuller D.J., Henriksen A., Smith A.T., Poulos T.L.;
 RT "Crystal structure of horseradish peroxidase C at 2.15-A resolution."; Nat. Struct. Biol. 4:1032-1038 (1997).
 RL [4]
 RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RP MEDLINE=98272679; PubMed=9609699;
 RA Henriksen A., Schuller D.J., Meno K., Welinder K.G., Smith A.T.,
 RT "Structural interactions between horseradish peroxidase C and the substrate benzohydroxamic acid determined by X-ray crystallography."; Biochemistry 37:8054-8060 (1998).
 RL [5]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RP Meno K., White C.G., Smith A.T., Gajhede M.;
 RT Submitted (DEC-1998) to the PDB data bank.
 RL Submitted (DEC-1998) to the PDB data bank.
 CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS, BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.
 CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -1- COFACTOR: Binds 1 prothene IX, 1 iron(III) ion and 2 calcium ions per subunit.
 CC -1- SIMILARITY: MONOMER.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. PLANT PEROXIDASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, M37156; AAA33377.1; ALT_SEQ.
 DR PIR, A00502; OPRHC.
 DR PIR, S00625; S00625.
 DR PDB, 1ATU; 04-FEB-98.
 DR PDB, 2ATU; 28-JAN-98.
 DR PDB, 3ATU; 23-DEC-98.
 DR GlycoSiteDB, P00433;
 DR InterPro, IPR002016; Peroxidase.
 DR Pfam, PF00141; Peroxidase; 1.
 DR PRINTS, PR00458; PEROXIDASE.
 DR PROSITE, PS00435; PEROXIDASE_1; 1.
 DR PROSITE, PS00436; PEROXIDASE_2; 1.
 DR Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;
 KM Multigene family; Signal; 3D-structure.
 FT SIGNAL 1 30
 FT CHAIN 1 338 PEROXIDASE CIA.
 FT PROPEP 339 353
 FT CA_BIND 73 73 1.
 FT CA_BIND 76 76 1 (VIA CARBONYL OXYGEN).
 FT CA_BIND 78 78 1 (VIA CARBONYL OXYGEN).
 FT CA_BIND 80 80 1.
 FT CA_BIND 82 82 1.
 FT CA_BIND 94 94 1.
 FT CA_BIND 201 201 2.
 FT CA_BIND 236 236 2.
 FT CA_BIND 239 239 2.
 FT CA_BIND 242 242 2.
 FT CA_BIND 244 244 2.
 FT CA_BIND 257 257 2.
 FT ACT_SITE 68 68
 FT ACT_SITE 72 72
 FT METAL 200 200
 FT MOD_RES 31 31
 FT DISULFID 41 121
 FT DISULFID 74 79
 FT DISULFID 127 331
 FT DISULFID 207 239
 FT CARBOHYD 43 43
 N-LINKED (GLCNAC. . .).
 DISTAL HISTIDINE.
 IRON (PROHEME IX AXIAL LIGAND).
 PYRROLIDONE CARBOXYLIC ACID.

FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .)
 SQ SEQUENCE 353 AA; 38825 MW; AC916C03C4A2AD27 CRC64;
 Query Match 0.8%; Score 9; DB 1; Length 353;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 680 VTLAGPS 687
 Db 139 VTLAGPS 146
 RESULT 3
 PERC_ARATH STANDARD; PRT; 354 AA.
 AC P24101;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neutral peroxidase C precursor (EC 1.11.1.7).
 GN PRXC OR AT3G49110 OR T2J13.50.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucoisid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91200671; PubMed=2016063;
 RA Intaprun C., Higashimura N., Yamamoto K., Okada N., Shimmyo A.,
 RA Takano M.;
 RT "Nucleotide sequences of two genomic DNAs encoding peroxidase of Arabidopsis thaliana."; Gene 98:237-241 (1991).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansgore W., Unseid M.,
 RA Fartman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA De Simone V., Cholsen N., Artiguenave F., Robert C., Broetier P.,
 RA Wincker P., Cattolico L., Weissendach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Topo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
 RA Reichelt J., Schafte M., Schoen O., Bargues M., Terol J., Clement J.,
 RA Navarro P., Collado C., Perez-Perez A., Outenweider B., Duchemin D.,
 RA Cooke R., Landie M., Berger-Liauro C., Purnelle B., Masny D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Montfort A., Argitrou A., Flores M., Liguori R., Vitale D.,
 RA Mannhaupt G., Haase D., Schoof H., Rued S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Uteback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Miltseher J., Sellers P., Gill J.E., Feldlym T.V.,
 RA Preuss D., Lin X., Niernm W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Ideawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Iwasawa M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana."; Nature 408:820-822 (2000).
 RL

```

CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,
CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD
CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE
CC DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.
CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC -1- COFACTOR: Binds 1 prothene IX and 1 iron(III) ion.
CC -1- TISSUE SPECIFICITY: ROOTS.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. PLANT PEROXIDASE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M58380; AAA32849.1; -
DR EMBL; AL132967; CAB61999.1; -
DR PIR; J00457; J00457.
DR HSSP; P00433; ZATU.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
KW Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;
KW Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 354
FT ACT SITE 69 69 NEUTRAL PEROXIDASE C.
FT ACT SITE 73 73 DISTAL HISTIDINE (BY SIMILARITY).
FT METAL 201 201 IRON (PROTOHEME IX AXIAL LIGAND).
FT DISULFID 42 122 BY SIMILARITY.
FT DISULFID 75 80 BY SIMILARITY.
FT DISULFID 128 332 BY SIMILARITY.
FT DISULFID 208 240 BY SIMILARITY.
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 354 AA; 38941 MW; 6D2BE3D53611724 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 680 VTLAGPS 687
Db 140 VTLAGPS 147

RESULT 4
TOLB_RICPR STANDARD; PRT; 443 AA.
ID TOLB_RICPR
AC Q9ZDM5;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLB protein precursor.
GN TOLB OR RP302.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OC NCBI_TaxID=782;
RX MEDLINE=99039495; PubMed=9823893;
RA Anderson S.G.E.; Zomrodipour A.; Andersson J.O.;
RA Sichteritz-Ponten T.; Alsmark U.C.M.; Podowcki R.M.; Naeslund A.K.;

```

```

RA Eriksen A.-S.; Winkler H.H.; Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria";
RL Nature 396:133-140(1998).
CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE TOLB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235271; CAA14763.1; -
DR EMBL; AJ235271; CAA14763.1; -
KW Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 18
FT CHAIN 19 443 TOLB PROTEIN.
SQ SEQUENCE 443 AA; 49206 MW; BE5BC2282A0F5593 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 443;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 PRGDIYAF 168
Db 340 PRGDIYAF 347

RESULT 5
DYHC_MOUSE STANDARD; PRT; 4644 AA.
ID DYHC_MOUSE
AC Q9JH14;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dynein heavy chain, cytosolic (DYHC) (Cytoplasmic dynein heavy chain).
GN DNCH1 OR DNCH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX Sasaki S.; Shionoya A.; Hirotsune S.;
RT "Complete cDNA sequence of murine cytoplasmic dynein heavy chain.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY004877; AAF91078.1; -
DR MGD; MG1103147; DNCH1.
DR InterPro; IPR004273; Dynein heavy.
DR Pfam; PF03028; Dynein heavy; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 48 69 COILED COIL (POTENTIAL).

```

FT DOMAIN 179 200 COILED COIL (POTENTIAL).
 FT DOMAIN 453 476 COILED COIL (POTENTIAL).
 FT DOMAIN 541 564 COILED COIL (POTENTIAL).
 FT DOMAIN 1169 1201 COILED COIL (POTENTIAL).
 FT DOMAIN 1229 1250 COILED COIL (POTENTIAL).
 FT DOMAIN 1355 1371 COILED COIL (POTENTIAL).
 FT DOMAIN 2012 2040 MICROTUBULE-BINDING (POTENTIAL).
 FT DOMAIN 3187 3273 COILED COIL (POTENTIAL).
 FT DOMAIN 3394 3498 COILED COIL (POTENTIAL).
 FT DOMAIN 3735 3798 COILED COIL (POTENTIAL).
 FT NP BIND 1904 1911 ATP (POTENTIAL).
 FT NP BIND 2222 2229 ATP (POTENTIAL).
 FT NP BIND 2593 2600 ATP (POTENTIAL).
 FT NP BIND 2935 2942 ATP (POTENTIAL).
 SQ SEQUENCE 4644 AA; 532021 MW; F55B4E15DD479E1B CRC64;

Query Match 0.8%; Score 8; DB 1; Length 4644;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SARVGRGT 18
 Db 720 SARVGRGT 727

RESULT 6
 ID_NER_BPDI0 STANDARD; PRT; 73 AA.
 AC P06903;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-binding protein Ner.
 GN NER.
 OS Bacteriophage D108.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC Mu-like viruses.
 OX NCBI_Taxid=10671;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86055744; PubMed=2998774;
 RA Tojias P.P., Dubow M.S.;
 RT "The cloning and characterization of the bacteriophage D108
 RT regulatory DNA-binding protein ner.";
 RL EMOB J. 4:3031-3037(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86232621; PubMed=3012481;
 RA Mizunuchi M., Weisberg R.A., Mizunuchi K.;
 RT "DNA sequence of the control region of phage D108: the N-terminal
 RT amino acid sequences of repressor and transposase are similar both in
 RT phage D108 and in its relative, phage M1.";
 RL Nucleic Acids Res. 14:3813-3825 (1986).
 CC -1- FUNCTION: BINDS TO THE REGION RIGHT AFTER THE START OF
 CC TRANSCRIPTION OF THE PROMOTER REGION OF THE EARLY GENES.
 CC -1- SIMILARITY: BELONGS TO THE NER FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M26291; AAA32206.1; -;
 CC DR EMBL; X03847; CAA27475.1; -;
 CC DR PIR; A24680; DNBPD8.
 CC DR PIR; S07931; S07931.
 CC DR HSSP; P06020; INEO.
 CC KW Transcription regulation; DNA-binding.
 CC KW DNA_BIND 28 47 H-T-H MOTIF (PROBABLE).

SQ SEQUENCE 73 AA; 8534 MW; 89BDFD38E1409025 CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 73;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 982 LGRSNHL 988
 Db 31 LGRSNHL 37

RESULT 7
 ID_LSM6_SCHPO STANDARD; PRT; 75 AA.
 AC Q9UTL1.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Probable U6 snRNA-associated Sm-like protein LSM6.
 GN SPAC2F3.17C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_Taxid=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymopiez B.,
 RA Wellens I., Vansteleels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambolt S., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakowski G.V., Useary D., Bartell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE SNRP SM PROTEINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; Z99165; CAB54975.1; -;
 CC DR InterPro; IPR001163; snRNP_Sm.
 CC DR Pfam; PF01423; Sm_1.
 CC KW Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;
 CC KW RNA-binding.
 CC SQ SEQUENCE 75 AA; 8336 MW; 20B1B1F2E380BCCF CRC64;

Query Match 0.7%; Score 7; DB 1; Length 75;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

237 GGNVLY 243
 |||||
 Db 63 GGNVLY 69

RESULT 8
 LSM6 HUMAN
 ID LSM6 HUMAN STANDARD; PRT; 80 AA.
 AC 094YB;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE U6 snRNA-associated Sm-like protein LSM6.
 GN LSM6.
 OS Homo sapiens (Human), and
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606, 10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Human;
 RX MEDLINE=9238196; PubMed=10369684;
 RA Salgado-Garrido J., Bragado-Nilsson E., Kandelis-Lewis S., Seraphin B.,
 RT "Sm and Sm-like proteins assemble in two related complexes of deep
 RT evolutionary origin."
 RL EMBO J. 18:3451-3462(1999).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC SPECIES=Human;
 RX MEDLINE=9452783; PubMed=10523320;
 RA Achsel T., Brahms H., Kastner B., Bachl A., Wilm M., Luehmann R.;
 RT "A doughnut-shaped heteromer of human Sm-like proteins binds to the
 RT 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in
 RT vitro."
 RL EMBO J. 18:5789-5802(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Mouse; STRAIN=C57BL/6J;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schirni L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barah G.,
 RA Blake J., Boffelli D., Botjunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RL Nature 409:685-690(2001).
 CC -1- FUNCTION: BINDS SPECIFICALLY TO THE 3'-TERMINAL U-TRACT OF U6
 CC SNRNA.
 CC -1- SUBUNIT: LSM SUBUNITS FORM A HETEROMER WITH A DOUGHNUT SHAPE.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potentially).
 CC -1- SIMILARITY: BELONGS TO THE SNRNP SM PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AJ238098; CAB45869.1; -
 CC DR EMBL, AF182292; AAD56230.1; -
 CC DR EMBL, AK019126; BAB31555.1; -
 CC DR InterPro, IPR001163; snRNP_Sm.
 CC DR Pfam, PF01423; Sm; 1.
 CC KW Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;
 CC RRNA-binding.
 CC SEQUENCE 80 AA; 9128 MW; 21167891FDE804F1 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 80;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

237 GGNVLY 243
 |||||
 Db 66 GGNVLY 72

RESULT 9
 URE3_ECOLI
 ID URE3_ECOLI STANDARD; PRT; 100 AA.
 AC 003282;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-1999 (Rel. 38, Last annotation update)
 DE Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).
 GN URE3.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OC Bacterioidia
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93194816; PubMed=8449894;
 RA D'Orazio S.E., Collins C.M.;
 RT "Characterization of a plasmid-encoded urease gene cluster found in
 RT members of the family Enterobacteriaceae."
 RL J. Bacteriol. 175:1860-1864(1993).
 CC -1- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
 CC -1- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, L03307; AAA24745.1; -
 CC DR PIR, B47090; B47090.
 CC DR HSPS, P18316; 1FMB.
 CC DR InterPro, IPR002026; Urease_gamma.
 CC DR Pfam, PF00547; urease_gamma; 1.
 CC DR ProDom, PD002319; Urease_gamma; 1.
 CC DR TIGRFAMs, TIGR00193; urease_gam; 1.
 CC KW Hydrolase; Plasmid.
 CC SEQUENCE 100 AA; 10983 MW; FBFD74F4C68B0805 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 100;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

960 SCAIMG 966
 |||||
 Db 40 SCAIMG 46

```

RESULT 10
URE3_PROVMI STANDARD; PRT; 100 AA.
AC P17088;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).
GN UREA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OC NCBI_TaxID=584;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=H14320;
RX MEDLINE=90078080; PubMed=2687233;
RA Jones B.D., Mobley H.L.T.;
RT "Proteus mirabilis urease: nucleotide sequence determination and
RT comparison with jack bean urease."
RL J. Bacteriol. 171:6414-6422(1989).
CC -1- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC -1- SUBUNIT: (ALPHA, BETA, GAMMA) (3) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31834; AAA2567.1; -.
DR PIR; B43719; B43719.
DR HSSP; P18316; 1FWB.
DR InterPro; IPR002026; Urease_gamma.
DR Pfam; PF00547; urease_gamma.1.
DR ProDom; PD002319; Urease_gamma.1.
DR TIGRFAMs; TIGR00193; urease_gam; 1.
DR HydroLase.
SQ SEQUENCE 100 AA; 10998 MW; 79E4A2182C1EF903 CRC64;
Query Match 0.7%; Score 7; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 960 SCAIMEG 966
DB 40 SCAIMEG 46

```

```

RESULT 11
URE3_PROVU STANDARD; PRT; 100 AA.
AC P16124;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).
GN UREA.
OS Proteus vulgaris.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OC NCBI_TaxID=585;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=DSM 30118;
RX MEDLINE=90215174; PubMed=2182382;
RA Moerdorf G., Kaltwasser H.;
RT "Cloning of the genes encoding urease from Proteus vulgaris and
RT sequencing of the structural genes."
RL FEMS Microbiol. Lett. 54:67-73(1990).
CC -1- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC -1- SUBUNIT: (ALPHA, BETA, GAMMA) (3) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51816; CA36113.1; -.
DR PIR; S08478; S08478.
DR HSSP; P18316; 1FWB.
DR InterPro; IPR002026; Urease_gamma.
DR Pfam; PF00547; urease_gamma.1.
DR ProDom; PD002319; Urease_gamma.1.
DR TIGRFAMs; TIGR00193; urease_gam; 1.
DR HydroLase.
SQ SEQUENCE 100 AA; 10955 MW; D9F413A9DA2C84B5 CRC64;
Query Match 0.7%; Score 7; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 960 SCAIMEG 966
DB 40 SCAIMEG 46

```

```

RESULT 12
DEF_RICCN STANDARD; PRT; 175 AA.
AC Q921Z1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptide deformylase (EC 3.5.1.88) (PDF) (polypeptide deformylase).
GN DEF OR RC0278.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia.
OX NCBI_TaxID=781;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barde V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."
RT Science 293:2093-2098(2001).
CC -1- FUNCTION: Removes the formyl group from the N-terminal Met of
CC newly synthesized proteins. Requires at least a dipeptide for an
CC efficient rate of reaction. N-terminal L-methionine is a
CC prerequisite for activity but the enzyme has broad specificity at
CC other positions (By similarity).
CC -1- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC methionyl peptide.
CC -1- COFACTOR: Binds 1 iron (II) ion (By similarity).
CC -1- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; AB008593; AA02816.1; -
 DR InterPro; IPR00181; pep_deformylase.
 DR Pfam; PF01327; pep_deformylase; 1.
 DR Prodom; PD003844; pep_deformylase; 1.
 DR Trifam; TIGR00079; pep_deformyl; 1.
 KW Protein biosynthesis; Hydrolyase; Iron; Complete proteome.
 FT METAL 99 99 IRON (BY SIMILARITY).
 FT METAL 141 141 IRON (BY SIMILARITY).
 FT ACT_SITE 142 142 BY SIMILARITY.
 FT METAL 145 145 IRON (BY SIMILARITY).
 SQ SEQUENCE 175 AA; 20279 MW; 93B050C6003CSA6D CRC64;

Query Match 0.7%; Score 7; DB 1; Length 175;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 892 ISLPEOR 898
 DB 100 ISLPEOR 106

RESULT 13
 ID Y773 METJA STANDARD; PRT; 175 AA.
 AC 058183;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M00773 precursor.
 GN M00773.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;

RA [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=6888087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerecsek A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weissbach K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073 (1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U67522; AAB98777.1; -
 DR TIGR; M00773; -
 KW Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 33 HYPOTHETICAL PROTEIN M00773.
 FT CHAIN 34 175
 SQ SEQUENCE 175 AA; 20246 MW; 63BE09A737E8F11 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 175;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 SFTSKGL 693
 DB 68 SFTSKGL 74

RESULT 14
 ID YCF4 ODOSI STANDARD; PRT; 181 AA.
 AC P49526;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Photosystem I assembly protein ycf4.
 GN YCF4.
 OS Odontella sinensis (Marine centric diatom).
 CC Chlorophyta; Stramenopiles; Bacillariophyta; Coscinodiscophyceae;
 CC Eukaryota; Eukaryotae; Eupodiscaceae; Eupodiscaceae; Odontella.
 OX NCBI_TaxID=2839;
 RA [1]

RP SEQUENCE FROM N.A.
 RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
 RT "The chloroplast genome of a chlorophyll a+c-containing alga,
 RT Odontella sinensis.";
 RL Plant Mol. Biol. Rep. 13:336-342 (1995).

CC -1- FUNCTION: Seems to be required for the assembly of the photosystem
 CC I complex (By similarity).
 CC -1- SUBCELLULAR LOCATION: Thylakoid membrane-associated (By
 CC similarity). BELONGS TO THE YCF4 FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z67753; CAA91709.1; -
 DR InterPro; IPR003359; PSI_Ycf4.
 DR Pfam; PF02392; Ycf4; 1.
 KW Photosynthesis; Thylakoid; Transmembrane; Chloroplast.
 FT TRANSMEM 19 41 POTENTIAL.
 FT TRANSMEM 61 83 POTENTIAL.
 SQ SEQUENCE 181 AA; 20547 MW; 663832A7DB554EBA CRC64;

Query Match 0.7%; Score 7; DB 1; Length 181;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 887 LCSGGIS 893
 DB 25 LCSGGIS 31

RESULT 15
 ID PMG2 ECOLI STANDARD; PRT; 215 AA.
 AC P36942;

DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable phosphoglycerate mutase 2 (EC 5.4.2.1) (Phosphoglyceromutase
 DE 2) (PGAM 2) (Bpg-dependent PGAM 2).
 GN GPMB OR B4395 OR Z5997 OR EC55353.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.

CC NCBI_TaxID=562, 83334;
 RA [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=KL2 / W3110;
 RX MEDLINE=93194823; PubMed=8449900;
 RA Skarstad K., Thoeny B., Hwang D.S., Kornberg A.;
 RA "A novel binding protein of the origin of the Escherichia coli

RT chromosome.";
 RL J. Biol. Chem. 268:5365-5370(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=95334362; PubMed=7610040;
 RA Burtland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=1126551;
 RA Berna N.T., Plunkett G. III, Burtland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsuno E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Sasakawa N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate
 CC = 3-phospho-D-glycerate + 2,3-diphosphoglycerate.
 CC -1- PATHWAY: Glycolysis.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE MUTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M97495; -; NOT_ANNOTATED_CDS.
 DR EMBL; U14003; AAA97291.1; -;
 DR EMBL; AE000509; AAC77348.1; -;
 DR EMBL; AE0005670; AAG59575.1; -;
 DR EMBL; AP002569; BAB38776.1; -;
 DR Ecogene; EG12164; gpmb.
 DR InterPro; IPR001345; Pg/BPGM_mutase.
 DR Pfam; PF00300; PGAM; 1.
 DR PROSITE; PS00175; PG_MUTASE; 1.
 KW Isomerase; Glycolysis; Multigene family; Complete proteome.
 FT ACT_SITE 9 FORMS THE PHOSPHOTITINE INTERMEDIATE
 FT ACT_SITE 9 (BY SIMILARITY).
 FT ACT_SITE 58 REQUIRED FOR BINDING CARBOXYL GROUP OF
 FT ACT_SITE 58 PHOSPHOGLYCERATES (BY SIMILARITY).
 FT ACT_SITE 150 BY SIMILARITY.
 FT ACT_SITE 35 O -> H (IN REF. 3).
 FT CONFLICT 148 V -> L (IN REF. 1).
 FT CONFLICT 148 V -> L (IN REF. 1).
 SO SEQUENCE 215 AA; 24065 MW; 3653DA0548B9E009 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 215;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Gy 722 RIBPGES 728
 Db 115 RIBPGES 121

RESULT 16
 ID DHSB_RECASM STANDARD; PRT; 239 AA.
 AC P80480;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Succinate dehydrogenase [ubiquinone] iron-sulfur protein (EC 1.3.5.1)
 DE (1P).
 DE SDH2 OR SDHB.
 OS Reclinomonas americana.
 OG Mitochondrion.
 OC Eukaryota; core jakobids; Reclinomonas.
 OX NCBI_TaxID=48483;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 50394;
 RX MEDLINE=96197311; PubMed=8637872;
 RA Burger G., Lang B.F., Reith M., Gray M.W.;
 RT "Genes encoding the same three subunits of respiratory complex II are
 RT present in the mitochondrial DNA of two phylogenetically distant
 RT eukaryotes.";
 RT Proc. Natl. Acad. Sci. U.S.A. 93:2328-2332(1996).
 CC -1- CATALYTIC ACTIVITY: Succinate + ubiquinone = fumarate + ubiquinol.
 CC -1- COFACTOR: BINDS THREE DIFFERENT IRON-SULFUR CLUSTERS: A 2FE-2S,
 CC A 3FE-4S AND A 4FE-4S (BY SIMILARITY).
 CC -1- PATHWAY: tricarboxylic acid cycle.
 CC -1- SUBUNIT: COMPOSED OF A 27 kDa IRON PROTEIN (1P) AND A 70 kDa
 CC FLAVOPROTEIN (FP).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF007261; AAD11913.1; -;
 DR HSSP; P00364; 1FCM.
 DR InterPro; IPR000564; 2FE2S_ferredoxin.
 DR InterPro; IPR001450; 4FE4S_ferredoxin.
 DR InterPro; IPR004489; DHSB.
 DR InterPro; IPR001041; Ferredoxin.
 DR Pfam; PF00111; fer2; 1.
 DR TIGRFAMs; TIGR00384; dhsb; 1.
 DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
 DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
 KW Tricarboxylic acid cycle; Iron-sulfur; Oxidoreductase; Mitochondrion;
 KW Electron transport; 3Fe-4S; 4Fe-4S.
 FT METAL 61 IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
 FT METAL 66 IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
 FT METAL 69 IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
 FT METAL 81 IRON-SULFUR S2 (3FE-4S) (BY SIMILARITY).
 FT METAL 151 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
 FT METAL 154 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
 FT METAL 157 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
 FT METAL 161 IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
 FT METAL 208 IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
 FT METAL 214 IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
 FT METAL 218 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
 SO SEQUENCE 239 AA; 27828 MW; 781C6F0D5375B CRC64;

Query Match 0.7%; Score 7; DB 1; Length 239;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Gy 220 WMTTEK 226
 Db 123 WMTTEK 129

RESULT 17
ID IF6_SCHPO STANDARD; PRT; 244 AA.
AC 094476;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 6 (eIF-6).
GN IF6 OR SPC1919.09.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsals K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,
RA Mooney P., Moulé S., Mungall K., Murphy L., Nblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Shelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicikert G., Aert R., Robben J., Grymptre B.,
RA Wolters I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filiz C., Holzer E., Moesli D., Hilbert H.,
RA Bozrym P., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Beyer P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leclaire V., Mottier S.,
RA Gallberg F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Roche M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC - FUNCTION: BINDS TO THE 60S RIBOSOMAL SUBUNIT AND PREVENTS ITS
CC ASSOCIATION WITH THE 40S RIBOSOMAL SUBUNIT TO FORM THE 80S
CC INITIATION COMPLEX (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE EIF-6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL035075; CAA22640.1; -
DR HSSP; Q12522; 1662.
DR InterPro; IPR002769; eIF6.
DR Pfam; PF01912; eIF6; 1.
DR ProDom; PD006880; eIF6; 1.
KW Initiation factor; protein biosynthesis.
SQ SEQUENCE 244 AA; 26232 MW; 388F52C670F25A3C CRC64;

Query Match 0.7%; Score 7; DB 1; Length 244;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

81 SLDPVK 87
|||||||

Db 87 SLDPVK 93
RESULT 18
ID TRY7_ANOGA STANDARD; PRT; 267 AA.
AC P35071;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin 7 precursor (EC 3.4.21.4).
GN TRY7.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Snakoko; TISSUE=Midgut;
RX MEDLINE=93327779; PubMed=8335004;
RA Mueller H.M., Crampston J.M., della Torre A., Sindén R., Crisanti A.,
RT "Members of a trypsin gene family in Anopheles gambiae are induced in
RT the gut by blood meal.";
RL EMO U. 12:2891-2900(1993).
CC - FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD
CC MEAL.
CC - CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.
CC - SUBCELLULAR LOCATION: Extracellular.
CC - TISSUE SPECIFICITY: MIDGUT.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z22930; CAA80516.1; -
DR PIR; S40006; S40006.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.112; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; TRY1_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Signal; Zymogen; Digestion;
KM Multigene family.
FT FT SIGNAL 1 18
FT PROPEP 19 41
FT CHAIN 42 267
FT ACT SITE 82 82
FT ACT SITE 127 127
FT ACT SITE 222 222
FT DISULFID 67 83
FT DISULFID 192 207
FT DISULFID 218 242
FT SITE 216 216
SQ SEQUENCE 267 AA; 28456 MW; 11015871257D8678 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 267;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

920 ALLTVL 926
|||||||
6 ALLTVL 12

```

RESULT 19
YC15_PRRHO STANDARD; PRT; 292 AA.
ID 058968;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical ABC transporter permease protein PH1215.
GN PH1215 OR PHBK039.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OC NCBI_Taxid=53953;
RN [1]
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa H., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaebacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC -1- SYSTEM PH1214/15/16. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION
CC OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. MALFG SUBFAMILY.
CC -----
CC CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000005; BAB30315.1; -
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF00528; BPD_transp.1.
DR KMW POSITIVE; PS00402; BPD_TRANSP_INN_MEMBER.1.
FT HYPOTHETICAL PROTEIN; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 106 126 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 261 281 POTENTIAL.
SQ SEQUENCE 292 AA; 33407 MW; 5B968812C0B89F68 CRC64;
Query Match 0.7%; Score 7; DB 1; Length 292;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 826 VPGSILL 832
Db 79 VPGSILL 85

```

```

DE processing protein 4).
GN RRP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RX MEDLINE=9605921; PubMed=9390555;
RA Mitchell P., Petfalaki E., Shevchenko A., Mann M., Tollervey D.;
RT "The exosome: a conserved eukaryotic RNA processing complex containing
RT multiple 3'--5' exonucleases."
RL Cell 91:457-466(1997).
RN [4]
RX SEQUENCE FROM N.A.
RN TISSUE=Lung;
RA Straube R.;
RN Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RX CHARACTERIZATION.
RX MEDLINE=96178774; PubMed=8600032;
RA Mitchell P., Petfalaki E., Tollervey D.;
RT "The 3' end of yeast 5.8S rRNA is generated by an exonuclease
RT processing mechanism."
RL Genes Dev. 10:502-513(1996).
RN RN CHARACTERIZATION.
RX MEDLINE=9936719; PubMed=10465791;
RA Allmann C., Petfalaki E., Podtelejnikov A., Mann M., Tollervey D.,
RA Mitchell P.;
RT "The yeast exosome and human PM-Scl are related complexes of 3'--5'
RT exonucleases."
RL Genes Dev. 13:2148-2158(1999).
CC -1- FUNCTION: BELONGS TO THE EXOSOME, A RNA PROCESSING COMPLEX, WHICH
CC IS AT LEAST INVOLVED IN THE 3' PROCESSING OF THE 7S PRE-RNA TO
CC THE MATURE 5.8S RNA. EXHIBITS A 3'-5' EXORIBONUCLEASE ACTIVITY
CC (BY SIMILARITY).
CC -1- SUBUNIT: COMPONENT OF THE EXOSOME MULTISUBUNIT RIBONUCLEASE COMPLEX
CC COMPOSED OF AT LEAST 11 PROTEINS: RRP4, RRP40, RRP41/SKI6, RRP42,
CC RRP43, RRP44/DIS3, PM/SCI-75, RRP46, CSL4 AND PM/SCI-100 (ONLY IN
CC THE NUCLEAR COMPLEX). ALSO ASSOCIATED WITH THE GTPASE RAN (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 S1 MOTIF DOMAIN.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO THE
CC SKIPPING OF AN EXON.
CC -----
CC CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U07561; AAB60392.1; ALT_SEQ.
DR EMBL: AK001916; BAB1977.1; -
DR EMBL: AK022460; BAB14043.1; -
DR EMBL: BC000747; AAB00747.1; -

```

DR MTM: 602238; -;
 DR InterPro: IPR003029; SL.
 DR SMART: SM00316; SL; 1.
 DR Exosome; Hydrolase; Nuclease; Exonuclease; rRNA processing;
 KM Nucleic protein; RNA-binding.
 FT DOMAIN 79 159 S1 MOTIF.
 SO SEQUENCE 293 AA; 32789 MW; 882033F50791643F CRC64;

Query Match 0.7%; Score 7; DB 1; Length 293;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 PVSILADR 765
 Db 225 PVSILADR 231

RESULT 21
 ID DPCC ECOLI STANDARD; PRT; 300 AA.
 AC P37315;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dipeptide transport system permease protein dpcc.
 GN DPCC OR B3542 OR Z4959 OR EC54422.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MM500;
 RC MEDLINE=95231288; PubMed=7536291;
 RA Abouhamed W.N., Manson M.D.;
 RT "The dipeptide permease of Escherichia coli closely resembles other
 RT bacterial transport systems and shows growth-phase-dependent
 RT expression".
 RL Mol. Microbiol. 14:1077-1092(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RC MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes."
 RL Nucleic Acids Res. 22:2576-2586(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RC MEDLINE=21074935; PubMed=11206551;
 RA Berna N.T., Plunkett G. III, Burland V., Mau B., Glaener J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potanous K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RC MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC -1- FOR DIPEPTIDES, PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF

CC THE SUBSTRATE ACROSS THE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM PERMEASE FAMILY. OPEBC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----

DR EMBL; L08399; AAA23704.1; -;
 DR EMBL; U00039; AAB18520.1; -;
 DR EMBL; AE000431; AAC76567.1; -;
 DR EMBL; AE005580; AAG58686.1; -;
 DR EMBL; AP002565; BAB37845.1; -;
 DR Ecogene; EG12626; dpcc.
 DR InterPro; IPR000515; BPD_transp.
 DR Pfam; PF00528; BPD_transp. 1.
 DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER, FALSE NEG.
 KM Transport; Peptide transport; Transmembrane; Inner membrane;
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 102 122 POTENTIAL.
 FT TRANSMEM 137 157 POTENTIAL.
 FT TRANSMEM 207 227 POTENTIAL.
 FT TRANSMEM 231 251 POTENTIAL.
 FT TRANSMEM 266 286 POTENTIAL.
 SO SEQUENCE 300 AA; 32308 MW; C4DD7BD82286E62 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 300;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 920 AILLTVL 926
 Db 275 AILLTVL 281

RESULT 22
 ID STBA ECOLI STANDARD; PRT; 320 AA.
 AC P11904; P10028;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-1999 (Rel. 38, Last annotation update)
 DE Protein Stba (Para locus 36 kDa protein).
 GN STBA.
 OS Escherichia coli.
 OG Plasmid IncFII NR1, and Plasmid IncFII R1.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KP245; PLASMID=IncFII NR1;
 RC MEDLINE=89011976; PubMed=3172224;
 RA Tabuchi A., Min Y.-N., Kim C.K., Fan Y.-L., Womble D.D., Rownd R.H.;
 RT "Genetic organization and nucleotide sequence of the stability locus
 RT of IncFII plasmid NR1.";
 RL J. Mol. Biol. 202:511-525(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC PLASMID=IncFII R1;
 RC MEDLINE=87060986; PubMed=3023637;
 RA Gerdes K., Molin S.;
 RT "Partitioning of plasmid R1. Structural and functional analysis of
 RT the para locus.";
 RL J. Mol. Biol. 190:269-279(1986).
 RN [3]
 RP REVISIONS TO 39 AND 60.

RC PLASMID=Ineffi R1;
 RA Gerdas K.;
 RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN IS THOUGHT TO BE INVOLVED IN THE
 CC CONTROL OF PLASMID PARTITION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X13777; CAA31264.1; -;
 DR EMBL: X04268; CAA27818.1; -;
 DR PIR: S01774; S01774.
 DR PIR: A24920; A24920.
 KM Plasmid; Plasmid partition.
 SQ SEQUENCE 320 AA; 35765 MW; 52B3239C0A567312 CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 320;
 Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 386 FFKTNNS 392
 Db 301 FFKTNNS 307
 RESULT 23
 YA0D SCHPO STANDARD; PRT; 346 AA.
 ID YA0D SCHPO
 AC Q09681;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative glycosyl transferase CSH10.13c in chromosome I (EC 2.-.-.-).
 GN SPAC5H10.13c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
 RA Brooke K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gertles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voicakeert G., Aert R., Robben J., Gymnopoulos B.,
 RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moser D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Drenth S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";

RL Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 34.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z49811; CAA89963.1; -;
 KM Hypothetical protein; Transferase; Glycosyltransferase; Transmembrane;
 KM Signal-anchor; Golgi stack; Glycoprotein. (POTENTIAL).
 FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 12 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT FT DOMAIN 33 346 LUMENAL (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 224 224 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 346 AA; 39107 MW; 1AD26CDDFAC0D3E CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 346;
 Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 556 TTFHEHS 562
 Db 81 TTFHEHS 87
 RESULT 24
 AROG ECOLI STANDARD; PRT; 350 AA.
 ID AROG ECOLI
 AC P00886;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phospho-2-dehydro-3-deoxyheptonate aldolase, phe-sensitive
 DE (EC 4.1.2.15) (Phospho-2-keto-3-deoxyheptonate aldolase) (DAP
 DE synthetase) (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase).
 GN AROG OR B0754 OR Z0324 OR ECS0782.
 OS Escherichia coli, and
 OS Escherichia coli 0157:H7.
 OC Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=82274236; PubMed=6125934;
 RA Davies W.D., Davidson B.E.;
 RT "The nucleotide sequence of aroG, the gene for 3-deoxy-D-
 RT arabinoheptulosonate-7-phosphate synthetase (phe) in Escherichia coli
 RT K12.";
 RT Nucleic Acids Res. 10:4045-4058(1982).
 RL [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plummer G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RL [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kaehimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.U., Davis N.W., Lam A., Dimalanta E.T., Potamocous K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 1-12.
RC STRAIN=K12 / EWG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:11259-1131(1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=99354419; PubMed=10425687;
RA Shumlin I.A., Kretsinger R.H., Baerle R.H.;
RT "Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-
RT heptulosonate-7-phosphate synthase from *Escherichia coli*.";
RL Structure 7:865-875(1999).
CC -1- FUNCTION: STEROSPECIFIC CONDENSATION OF PHOSPHOENOLPYRUVATE (PEP)
CC AND D-ERYTHROSE-4-PHOSPHATE (E4P) GIVING RISE TO 3-DEOXY-D-
CC ARABINO-HEPTULOSONATE-7-PHOSPHATE (DAHP).
CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-arabino-heptonate 7-
CC phosphate + phosphate = phosphoenolpyruvate + D-erythrose 4-
CC phosphate + H(2)O.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC first step.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- MISCELLANEOUS: THERE ARE 3 DAHP SYNTHASES, AROG IS
CC FEEDBACK-INHIBITED BY PHE. THE OTHER 2 DAHP SYNTHASES ARE TYR- AND
CC TRP-SENSITIVE, RESPECTIVELY.
CC -1- SIMILARITY: BELONGS TO CLASS-I DAHP SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J01591; AAA23492.1; -
CC EMBL: AB000178; AAC73841.1; -
CC DR EMBL: D90714; BAA35416.1; -
CC DR EMBL: AB005253; AAG5083.1; -
CC DR EMBL: AP002553; BAB34205.1; -

DR PIR: A01106; ADECHP.
DR PDB: 1QR7; 31-ARG-99.
DR SWISS-2DPAGE: P00886; COLI.
DR EcoGene: EG10079; arog.
DR InterPro: IPR001785; arog.
DR Pfam: PF00793; DAHP synth.1; 1.
DR ProDom: PD005060; DAHP synth.1.
DR TrRfam: TRF00034; aroFeh; 1.
DR Aromatic amino acid biosynthesis; Lyase; 3D-structure;
DR Complete proteome.
SQ SEQUENCE 350 AA; 38009 MW; 7477D361962E8710 CRC64;
Qy 97 AGEFLDM 103
Db 141 AGEFLDM 147
Query Match 0.7%; Score 7; DB 1; Length 350;
Best Local Similarity 100.0%; Pred. No.45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT 25
AROG HAEIN STANDARD; PRT; 362 AA.
ID AROG HAEIN
AC P4303;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospho-2-dehydro-3-deoxyheptonate aldolase (EC 4.1.2.15) (Phospho-
DE 2-keto-3-deoxyheptonate aldolase) (DAHP synthetase) (3-deoxy-
DE D-arabino-heptulosonate 7-phosphate synthase).
GN AROG OR H11547.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleisemann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-L., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddlow E., Cotton M.D.,
RA Utechtack T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "whole-genome random sequencing and assembly of *Haemophilus influenzae*
RT R.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: STEROSPECIFIC CONDENSATION OF PHOSPHOENOLPYRUVATE (PEP)
CC AND D-ERYTHROSE-4-PHOSPHATE (E4P) GIVING RISE TO 3-DEOXY-D-
CC ARABINO-HEPTULOSONATE-7-PHOSPHATE (DAHP).
CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-arabino-heptonate 7-
CC phosphate + phosphate = phosphoenolpyruvate + D-erythrose 4-
CC phosphate + H(2)O.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC first step.
CC -1- SIMILARITY: BELONGS TO CLASS-I DAHP SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32830; AAC23197.1; -
CC CC HSSP: P00886; 1QR7.
CC DR TIGR: H11547; -

CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
 CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
 CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
 CC -1- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X79404; CAAS5941.1; -
 DR PIR; S45051; S45051.
 DR InterPro; IPR000217; Tubulin.
 DR InterPro; IPR003008; Tubulin_Pfez.
 DR Pfam; PF00091; tubulin_1.
 DR PRINTS; PR01161; TUBULIN.
 DR PROSITE; PS00227; TUBULIN; 1.
 DR Microtubules; GTP-binding; Multigene family.
 KW NP BIND 142 148 GTP (POTENTIAL).
 FT SITE 451 451 INVOLVED IN POLYMERIZATION.
 SQ SEQUENCE 451 AA; 50254 MW; CC058D548A6946F5 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 451;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 964 MEGEDVE 970
 Db 442 MEGEDVE 448

RESULT 29
 RADA_LISMO STANDARD; PRT; 457 AA.
 AC Q48761; O86063;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA repair protein rada homolog (DNA repair protein sms homolog).
 GN RADA OR SMS OR SARII OR LM00233.
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_Taxid=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LO28 / Serovar 1/2c;
 RA Rouquette C.E., Tascou R., Pellegrini E., Bolla J., Berche P.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2175 / Serovar 4b;
 RA Truong T.K., Katharion S.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-e / Serovar 1/2a;
 RA MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Blocker H., Brandt P., Chakraborty T.,
 RA Chabib A., Chetouani F., Couve E., de Darvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Eutian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Knäuper G.,
 RA Madueno E., Maicorum A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstrek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vaquero-Boland J.-A., Voss H., Weiland J., Cossart P.,
 RN "Comparative genomics of Listeria species.";

RL Science 294:849-852(2001).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE REPAIR OF ENDOGENOUS ALKYLATION
 CC DAMAGE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RECA FAMILY. RADA SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U40604; AAC44447.1; -
 DR EMBL; AF083254; AAC33293.1; -
 DR EMBL; AF591974; CAD00760.1; -
 DR Listlist; LMO00233;
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR001984; AAA_ATPase.
 DR InterPro; IPR004504; Sms.
 DR PRINTS; PR00830; ENDOLAPTASE.
 DR ProDom; PD000229; RecA; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00416; Sms; 1.
 DR PROSITE; PS50162; RECA_2; 1.
 KW DNA repair; ATP-binding; Zinc-finger; DNA-binding; Complete proteome.
 FT NP BIND 12 29
 FT NP BIND 97 104
 FT CONFLICT 37 37
 FT CONFLICT 73 73
 FT CONFLICT 75 75
 FT CONFLICT 86 86
 FT CONFLICT 132 132
 FT CONFLICT 181 181
 FT CONFLICT 197 197
 FT CONFLICT 226 227
 FT CONFLICT 244 244
 FT CONFLICT 256 256
 FT CONFLICT 260 261
 FT CONFLICT 290 290
 FT CONFLICT 293 293
 FT CONFLICT 299 317

Query Match 0.7%; Score 7; DB 1; Length 457;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 ECTATLM 178
 Db 195 ECTATLM 201

RESULT 30
 CABI_METUA STANDARD; PRT; 482 AA.
 AC Q58773;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Carboxyl-phosphate synthase large chain, N-terminal section (EC 6.3.5.5) (Carboxyl-phosphate synthetase ammonia chain).
 GN CARI OR M1378.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 OX NCBI_Taxid=2190;
 RN [1]

SEQUENCE FROM N.A. / DSM 2661 / ATCC 43067;
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kierulff R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kistner E.F., Weissbach K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 RA Uettermann T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii." Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + L-glutamate + carbamoyl phosphate.
 CC -1- COFACTOR: Binds three manganese ions (By similarity).
 CC -1- PATHWAY: Arginine biosynthesis.
 CC -1- SUBUNIT: Pyrimidine biosynthesis; first step.
 CC -1- PROMOTES: Composed of two chains; the small (or glutamine) chain
 CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate.
 CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY, N-TERMINAL SECTION.
 CC -1- CAUTION: Sequence of carb is split into two genes in M.jannaschii
 CC (MJ1378 and MJ1381).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U67577; AAB9385.1; -.
 CC HSSP; P00968; IUDB.
 CC TIGR; MJ1378; -.
 DR InterPro; IPR005483; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.
 DR InterPro; IPR005481; CPase_L_N.
 DR Pfam; PF00289; CPase_L chain; 1.
 DR Pfam; PF02786; CPase_L_D2; 1.
 DR Pfam; PF02787; CPase_L_D3; 1.
 DR PRINTS; PR00098; CPASE.
 DR PROSITE; PS00866; CPASE_1; 1.
 DR PROSITE; PS00867; CPASE_2; 1.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; ATP-binding;
 KW Arginase; Complete proteome.
 FT DOMAIN 1 398 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT NP_BIND 150 207
 FT NP_BIND 299 349 ATP (POTENTIAL).
 FT METAL 281 281 MANGANESE 1 (BY SIMILARITY).
 FT METAL 295 295 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 297 297 MANGANESE 2 (BY SIMILARITY).
 SQ SEQUENCE 482 AA; 53608 MW; ABAABE570B23D96 CRC64;
 Query Match 0.7%; Score 7.7; DB 1; Length 482;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Penton protein (Viron component III) (Penton base protein).
 GN PIII.
 OS Avian adenovirus galli (strain Phelps) (Fowl adenovirus 1) (CELO).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenovirus.
 OX NCBI_TaxID=10553;
 RN [1]
 RP SEQUENCE FROM N.A. / PubMed=6827769;
 RX MEDLINE=96186720; PubMed=6827769;
 RA Chlocca S., Kurzbauner R., Schaffner G., Baker A., Mautner V.,
 RA Cotten M.;
 RT "The complete DNA sequence and genomic organization of the avian
 RT adenovirus CELO." J. Virol. 70:2939-2949(1996).
 RL J. Virol. 70:2939-2949(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U46933; AAC54908.1; -.
 CC InterPro; IPR002605; Adeno_Penton_B.
 DR Pfam; PF01686; Adeno_Penton_B; 1.
 KW Late protein.
 SQ SEQUENCE 515 AA; 56722 MW; 90C389ACD686C6AC CRC64;
 Query Match 0.7%; Score 7.1; DB 1; Length 515;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

720 DLRIPG 726
 DB 181 DLRIPG 187

RESULT 32
 RO60 HUMAN STANDARD; PRT; 538 AA.
 ID RO60 HUMAN
 AC P10155; Q92787;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 60-kDa SS-A/Ro ribonucleoprotein (60 kDa Ro protein) (60 kDa
 DE ribonucleoprotein Ro) (RoNP) (Ro 60 kDa autoantigen) (Sjogren
 DE syndrome type A antigen) (SS-A) (Sjogren syndrome antigen A2).
 GN SS-A2 OR RO60.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. / PubMed=3200833;
 RX MEDLINE=89071722; PubMed=3200833;
 RA Deutscher S.L., Harley J.B., Keene J.D.;
 RT "Molecular analysis of the 60-kDa human Ro ribonucleoprotein.";
 RT Proc. Natl. Acad. Sci. U.S.A. 85:9479-9483(1988).
 RN [2]
 RP SEQUENCE FROM N.A. / PubMed=2649513;
 RX MEDLINE=89198084; PubMed=2649513;
 RA Ben-Cherit E., Gandy B.J., Tan E.M., Sullivan K.F.;
 RT "Isolation and characterization of a cDNA clone encoding the 60-kD
 RT component of the human SS-A/Ro ribonucleoprotein autoantigen.";
 RT J. Clin. Invest. 83:1284-1292(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RA Buyon J.P., DiDonato F., Tseng C.E., Raehbaum W., Morris A.,
 RA Hamel J.C., Chan E.K.L.;
 RT "Identification and characterization of an alternative mRNA transcript
 RT of the 60-kD SS-A/Ro ribonucleoprotein encoding the N-terminal RNA
 RT binding domain alone." Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.

```

CC -1- FUNCTION: RNA-BINDING PROTEIN THAT BINDS TO SEVERAL SMALL
CC CYTOPLASMIC RNA MOLECULES KNOWN AS Y RNAS. MAY STABILIZE THESE
CC RNAS FROM DEGRADATION.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM/602; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DISEASE: SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS
CC OBTAIN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR
CC RO PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.
CC -1- SIMILARITY: BELONGS TO THE RO 60 KDA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04137; AAA35493.1; -
DR EMBL; M25077; AAA35532.1; -
DR EMBL; U44388; AAB81552.1; ALT_TERM.
DR EMBL; U44388; AAB81553.1; -
DR PIR; A31760; A31760.
DR PIR; A30596; A30596.
DR Genew; HGNC:11313; SSA2.
DR MIM; 600063; -
DR MIM; 234700; -
DR InterPro; IPR000504; RNA_rec_mot.
DR Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen;
KW Alternative splicing.
FT VARSPIC 195 205 LATVTKYITKG -> KHKIFIGKKG (IN SHORT
FT ISIFORM)
FT VARSPIC 206 538 MISSING (IN SHORT ISIFORM).
FT CONFLICT 239 239 K -> R (IN REF. 2).
FT CONFLICT 515 538 GMLDMCGFDTGALDVIRNFTLDMV -> ALQNTLINKSP
FT (IN REF. 2).
SQ SEQUENCE 538 AA; 60642 MW; B9C58AF3248DC48C CRC64;

Query Match 0.7%; Score 7; DB 1; Length 538;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 VTTDMTL 775
Db 424 VTTDMTL 430

RESULT 33
RO60 MOUSE STANDARD; PRT; 538 AA.
ID RO60 MOUSE
AC 008458; O9OYD8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60-kDa SS-A/Ro ribonucleoprotein (60 kDa Ro protein) (60 kDa
DE ribonucleoprotein Ro) (RoRNP).
GN SSA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=9726462; PubMed=9112230;
RX Wang D., Buynon J.P., Chan E.K.L.;
RA "Cloning and expression of mouse 60 kDa ribonucleoprotein SS-A/Ro.",
RT Mol. Biol. Rep. 23:205-210 (1996).
RL [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RA Kaufman K.M., Farris A.D., Gross J.K., Harley J.B.;

```

```

RT "Characterization of the murine 60 kD Ro gene: genomic sequence,
RT organization, and chromosomal localization.";
RT Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
RL [3]
RN RP SEQUENCE OF 82-538 FROM N.A.
RA Scottfield R.H., Kurten B.T., Kaufman K.M., James J.A., Baber U.,
RA Harley J.B.;
RT "Immunization of mice with human 60 kD Ro peptides results in epitope
RT spreading if the peptides are highly homologous between man and
RT mouse.";
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RNA-BINDING PROTEIN THAT BINDS TO SEVERAL SMALL
CC CYTOPLASMIC RNA MOLECULES KNOWN AS Y RNAS. MAY STABILIZE THESE
CC RNAS FROM DEGRADATION.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- TISSUE SPECIFICITY: HIGHEST IN BRAIN, FOLLOWED BY LUNG, MUSCLE,
CC KIDNEY AND HEART. LOWER LEVELS ARE FOUND IN TESTIS, LIVER AND
CC SPLEEN.
CC -1- SIMILARITY: BELONGS TO THE RO 60 KDA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U66843; AAC53142.1; -
DR EMBL; AF065398; AAF19049.1; -
DR EMBL; AF042139; AAC15667.1; -
DR MGI; MGI:106652; Ssa2.
KW Ribonucleoprotein; RNA-binding.
FT CONFLICT 32 33 D -> V (IN REF. 2).
FT CONFLICT 458 458 D -> G (IN REF. 2).
FT CONFLICT 465 465 V -> I (IN REF. 2).
SQ SEQUENCE 538 AA; 60123 MW; 7850DA35D1726BDA CRC64;

Query Match 0.7%; Score 7; DB 1; Length 538;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 VTTDMTL 775
Db 424 VTTDMTL 430

RESULT 34
GV7 XENLA STANDARD; PRT; 591 AA.
ID GV7 XENLA
AC P20398;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last annotation update)
DE Developmental protein xlgv7.
GN GV7.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=89252831; PubMed=2721962;
RX Miller M., Kloc M., Reddy B., Eastman E., Dreyer C., Beckin L.;
RA "xlgv7: a maternal gene product localized in nuclei of the central
RT nervous system in xenopus laevis.";
RL Genes Dev. 3:572-583 (1989).
RN RP SEQUENCE FROM N.A.
CC -1- FUNCTION: MATERNAL PROTEIN THAT MAY HAVE MULTIPLE FUNCTIONS
CC DURING DEVELOPMENT, ONE OF WHICH MAY BE ASSOCIATED WITH THE
CC DEVELOPMENT AND MAINTENANCE OF THE CENTRAL NERVOUS SYSTEM.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (FOLLOWING OOCYTE MATURATION),
CC THEN NUCLEAR (BLASTULA/GASTRULA STAGE).

```

CC -1- TISSUE SPECIFICITY: IN XENOPUS OOCYTE, IN THE CENTRAL NERVOUS
 CC SYSTEM CELLS OF STOMACH AND ADULT FROGS, AND TRANSIENTLY IN
 CC EPITHELIAL CELLS OF STOMACH AND GUT OF TADPOLES.
 CC -1- DEVELOPMENTAL STAGE: NEURULA STAGE AND IN ADULT BRAIN.
 CC -1- PPM: MAY BE PHOSPHORYLATED DURING OOCYTE MATURATION.
 CC -1- SIMILARITY: CLOSELY RELATED OR IDENTICAL TO NUCLEAR PROTEIN 7.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; X15627; CA33637.1; --
 CC PIR; A30098; A30098.
 CC InterPro; IPR001230; Prenyl site.
 CC Nuclear protein; Developmental protein; DNA-binding; Phosphorylation;
 CC ATP-binding; Brain; Differentiation; Repeat.
 CC NP_BIND 120 129 ATP (POTENTIAL).
 CC DOMAIN 579 583 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC DOMAIN 101 437 HIS-RICH AND REPEAT-RICH DOMAIN, POSSIBLE
 CC REPEAT 171 174 DNA-BINDING REGION.
 CC REPEAT 183 186
 CC REPEAT 224 227
 CC REPEAT 234 237
 CC SEQUENCE 591 AA; 66174 MW; 9F3364CE52B3B540 CRC64;

 Query Match 0.7%; Score 7f DB 1; Length 591;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 739 COAVIIP 745
 DB 509 COAVIIP 515

 RESULT 35
 MTHR_SCHPO STANDARD; PRT; 603 AA.
 ID MTHR_SCHPO
 AC Q10258;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable methyltetrahydrofolate reductase 1 (EC 1.5.1.20).
 GN SPACS6FR.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 OX RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones W., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moutre S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Shears S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grynolprez B.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Wellert Auer S.,
 RA Gabel K., Fuchs W., Fritze C., Holzer E., Moesel D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
 RA Dominguez A., Revela J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipakoveki G.V., Usero D., Barrett B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC -1- CATALYTIC ACTIVITY: 5-methyltetrahydrofolate + NADP(+) = 5,10-
 CC methylenetetrahydrofolate + NADPH.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- PATHWAY: Folate metabolism.
 CC -1- SIMILARITY: BELONGS TO THE METHYLENETERAHDROFOLATE REDUCTASE
 CC (EC 1.5.1.20/EC 1.7.99.5) FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; Z69728; CA93581.1; --
 CC HSSP; P00394; 1B5T.
 CC InterPro; IPR004621; Fadh2_euk.
 CC InterPro; IPR003117; Methylotrof_redctse.
 CC Pfam; PF02219; MTHR; 1.
 CC TRIGRAMS; TIGR00677; fadh2_euk; 1.
 CC Oxidoreductase; Flavoprotein; FAD; NADP.
 CC SEQUENCE 603 AA; 69012 MW; 38519FEB793D331F CRC64;

 Query Match 0.7%; Score 7f DB 1; Length 603;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 127 WDELPHG 133
 DB 342 WDELPHG 348

 RESULT 36
 CAB2_METUA STANDARD; PRT; 618 AA.
 ID CAB2_METUA
 AC Q58776;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Carbanoyl-phosphate synthase large chain, C-terminal section (EC
 DE 6.3.5.5) (Carbanoyl-phosphate synthetase ammonia chain).
 GN CAB2 OR MJ1381.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 OX RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96637999; PubMed=8688087;
 RA Butt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sulten G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utechback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii."
 RL Science 273:1058-1073(1996).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +

CC phosphate + L-glutamate + carbamoyl phosphate.
 CC -1- COFACTOR: Binds three manganese ions (By similarity).
 CC -1- PATHWAY: Arginine biosynthesis.
 CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY. C-TERMINAL SECTION.
 CC -1- CAUTION: Sequence of carb is split into two genes in M.jannaschii (MJ1378 and MJ1381).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U65778; AAB9391.1; -
 DR HSSP: P00968; 1BXK.
 DR TIGR: MJ1381; -
 DR InterPro: IPR005483; CPase_L.
 DR InterPro: IPR005479; CPase_L_D2.
 DR InterPro: IPR005480; CPase_L_D3.
 DR InterPro: IPR005481; CPase_L_N.
 DR InterPro: IPR004362; MGS_Like.
 DR Pfam: PF00289; CPase_L_Chain; 1.
 DR Pfam: PF02786; CPase_L_D2; 2.
 DR Pfam: PF02787; CPase_L_D3; 1.
 DR Pfam: PF02142; MGS; 1.
 DR PRINTS: PR00098; CPASE.
 DR PROSITE: PS00866; CPASE_1; 1.
 DR PROSITE: PS00867; CPASE_2; 1.
 DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; ATP-binding;
 KM Manganese; Complete proteome.
 FT DOMAIN 81 477
 FT DOMAIN 478 618
 FT METAL 358 358 ALLOSTERIC DOMAIN.
 FT METAL 370 370 MANGANESE 3 (BY SIMILARITY).
 FT METAL 370 370 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 618 AA; 68453 MW; 8C3D4D9C192A954E CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 618;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 137 LSANMEL 143
 Db 472 LSANMEL 478
 RESULT 37
 NCX3 RAT STANDARD; PRT; 624 AA.
 AC Q9EP00;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sodium/potassium/calcium exchanger 3 precursor (Na(+)/K(+)/Ca(2+) - exchange protein 3) (Fragment).
 GN SLC24A3 OR NCX3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=21303617; PubMed=11294860;
 RA Kraev A., Quedan B.D., Leach S., Li X.-F., Dong H., Winkeln R., Perizolo M., Cai X., Yang R., Phillips K.D., Lytton J.,
 RT "Molecular cloning of a third member of the potassium-dependent

.RT sodium-calcium exchanger gene family, NCX3.",
 RL J. Biol. Chem. 276:23161-23172 (2001).
 CC -1- FUNCTION: Transports one Ca(2+) and one K(+) in exchange for four Na(+). (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: Abundant in the brain. Expressed at low levels in the aorta, uterus and intestine.
 CC -1- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AY009158; AAG32680.1; -
 DR InterPro: IPR004481; K_NaCaexchng.
 DR InterPro: IPR004837; NaCa_Ex; 2.
 DR Pfam: PF01699; NaCa_Ex; 2.
 DR TIGRfam: TIGR00367; K_NaCaexchng-rel; 1.
 KM Sodium transport; Antiport; Symport; Calcium transport; Potassium transport; Sodium transport; Transmembrane; Glycoprotein; Signal; Repeat; Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 624
 FT DOMAIN 26 88
 FT TRANSMEM 89 109
 FT DOMAIN 110 133
 FT TRANSMEM 134 154
 FT DOMAIN 155 163
 FT TRANSMEM 164 184
 FT DOMAIN 185 191
 FT TRANSMEM 192 212
 FT DOMAIN 213 215
 FT TRANSMEM 216 236
 FT DOMAIN 237 465
 FT TRANSMEM 466 487
 FT DOMAIN 487 491
 FT TRANSMEM 492 512
 FT DOMAIN 513 530
 FT TRANSMEM 531 551
 FT DOMAIN 552 561
 FT TRANSMEM 562 582
 FT DOMAIN 583 596
 FT TRANSMEM 597 617
 FT DOMAIN 618 624
 FT REPEAT 130 170
 FT REPEAT 489 499
 FT CARBOHYD 52 52
 FT CARBOHYD 67 67
 SQ SEQUENCE 624 AA; 69388 MW; ADCAL75AF1B1B1141 CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 624;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 790 LGIPDVI 796
 Db 486 LGIPDVI 492
 RESULT 38
 NCX3_HUMAN STANDARD; PRT; 644 AA.
 AC Q9HC58; Q9BOY9; Q9BOL7; Q9BOY3; Q9H519;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sodium/potassium/calcium exchanger 3 precursor (Na(+)/K(+)/Ca(2+) - exchange protein 3).
 GN NCX3
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HeLa; TISSUE=Plasma;
 RX MEDLINE=11294860;
 RA Kraev A., Quedan B.D., Leach S., Li X.-F., Dong H., Winkeln R., Perizolo M., Cai X., Yang R., Phillips K.D., Lytton J.,
 RT "Molecular cloning of a third member of the potassium-dependent

ON SLC24A3 OR NCKX3.
 NS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain, and Skeletal muscle;
 RX MEDLINE=21303617; PubMed=11294880;
 RA Kreev A., Quedano B.D., Leach S., Li X.-F., Dong H., Winkfein R.,
 RA Perizzolo M., Cai X., Yang R., Phillips K.D., Lytton J.,
 RT "Molecular cloning of a third member of the potassium-dependent
 RT sodium-calcium exchanger gene family, NCKX3".
 RL J. Biol. Chem. 276:23161-23172 (2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 RA Bailey J., Bawlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Frazer A., French I., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lervahialho M.H., Leyerisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McCannachie L.J., McLeay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver S.A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Pratchalingam S.R., Plumb R.W., Ramey H.,
 RA Rice C.M., Ross M.T., Scott C.E., Selra H.K., Showstken R., Sims S.,
 RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulten J.B.,
 RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whiteaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:665-871(2001).
 CC -1- FUNCTION: Transports one Ca(2+) and one K(+) in exchange for four
 Na(+) (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Abundant in the brain. Expressed at low levels
 CC in the aorta, uterus and intestine.
 CC -1- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, AF288087; -; NOT ANNOTATED_CDS.
 DR EMBL, AF169257; AAC12968.2; -;
 DR EMBL, AL121830, CAC36055.1; -;
 DR EMBL, AL045647, CAC36055.1; -;
 DR EMBL, AL034425, CAC36051.1; -;
 DR EMBL, AL139429, CAC3155.1; -;
 DR GeneW, HGNC:10977, SLC24A3.
 DR InterPro, IPR004481; K_NaCaexchng.
 DR InterPro, IPR004837; NaCa_Exembh.
 DR Pfam, PF01699; Na_Ca_Ex; 4.
 DR TIGRFAMs, TIGR00367; K_NaCaexchng-rel; 1.
 KW Transport; Antipport; Symport; Calcium transport; Potassium transport;

KX	Sodium transport; Transmembrane; Glycoprotein; Signal; Repeat;
KW	Alternative splicing.
FT	SIGNAL 1 44 POTENTIAL.
FT	CHAIN 45 644 SODIUM/POTASSIUM/CALCIUM EXCHANGER 3.
FT	DOMAIN 45 107 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 108 128 POTENTIAL.
FT	DOMAIN 129 153 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 154 174 POTENTIAL.
FT	DOMAIN 175 182 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 183 203 POTENTIAL.
FT	DOMAIN 204 210 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 211 231 POTENTIAL.
FT	DOMAIN 232 234 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 235 255 POTENTIAL.
FT	DOMAIN 256 484 POTENTIAL.
FT	TRANSMEM 485 505 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 506 510 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 511 531 POTENTIAL.
FT	DOMAIN 532 549 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 550 570 POTENTIAL.
FT	DOMAIN 571 580 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 581 601 POTENTIAL.
FT	DOMAIN 602 615 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 616 636 EXTRACELLULAR (POTENTIAL).
FT	DOMAIN 637 644 ALPHA-1.
FT	REPEAT 149 189 ALPHA-2.
FT	REPEAT 518 549 POLY-GLU.
FT	DOMAIN 422 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 71 71 CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 401 441 MISSING (IN ISOFORM 2).
FT	VARSPLIC 442 442 S -> A (IN ISOFORM 2).
FT	CONFLICT 55 55 I -> V (IN REF. 1; FLAG12988 AND 2).
SO	SEQUENCE 644 AA; 72006 MW; E07EBB34EFA56D CRC64;

Q7	790 LGIPDV1 796	0.7%; Score 7; DB 1; Length 644;
		Best Local Similarity 100.0%; Pred. No. 78;
Db	505 LGIPDV1 511	Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

RESULT 39
NCKX_MOUSE STANDARD; PRT; 645 AA.
ID NCKX_MOUSE
AC Q99PD7; Q99PB8; Q99JR2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Sodium/potassium/calcium exchanger 3 precursor (Na(+)/K(+/Ca(2+)-
exchange protein 3).
GN SLG24A3 OR NCKX3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mamalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=129/SvJ, and CD-1; TISSUE=Embryonic stem cells, and Brain;
RC MEDLINE=21303617; PubMed=11294880;
RX Kraev A., Quehan B.D., Leach S.S., Li X.-F., Dong H., Winkfein R.,
RA Perizolo M., Cai X., Yang R., Phillips K.D., Lytton J.;
RT "molecular cloning of a third member of the potassium-dependent
RT sodium-calcium exchanger gene family," NCKX3."
RL J. Biol. Chem. 276:23161-23172(2001).
RN [2]
RN SEQUENCE OF 323-645 FROM N.A.
RP TISSUE=Breast tumor;
RC Strutsberg R.;
RA Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: Transports one Ca(2+) and one K(+) in exchange for four

```

CC Na(+) (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: Abundant in the brain. Highest levels found in
 CC selected thalamic nuclei, hippocampal CA1 neurons and in layer IV
 CC of the cerebral cortex.
 CC -1- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AF314821; AAC60049.1; -;
 CC DR EMBL; AF314822; AAC60050.1; -;
 CC DR EMBL; BC005742; AAH05742.1; -;
 CC DR MGI; MGI:2137513; SLC24a3.
 CC DR InterPro; IPR004481; K_NaCaexchng.
 CC DR Pfam; PF01699; Na_Ca_Ex_3.
 CC DR TIGR; TIGR00367; K_NaCaexchng-rel; 1.
 CC KW Transport; Antiport; Symport; Calcium transport; Potassium transport;
 CC Sodium transport; Transmembrane; Glycoprotein; Signal; Repeat.
 CC FT SIGNAL 1 43 POTENTIAL.
 CC FT CHAIN 1 43 SODIUM/POTASSIUM/CALCIUM EXCHANGER 3.
 CC FT DOMAIN 4 4 645 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 107 127 POTENTIAL.
 CC FT DOMAIN 128 151 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 152 172 POTENTIAL.
 CC FT DOMAIN 173 181 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 182 202 POTENTIAL.
 CC FT DOMAIN 203 209 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 210 234 POTENTIAL.
 CC FT DOMAIN 231 234 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 235 255 POTENTIAL.
 CC FT DOMAIN 256 286 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 287 307 POTENTIAL.
 CC FT DOMAIN 308 312 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 313 333 POTENTIAL.
 CC FT DOMAIN 334 351 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 352 372 POTENTIAL.
 CC FT DOMAIN 373 382 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 383 603 POTENTIAL.
 CC FT DOMAIN 604 617 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 618 638 POTENTIAL.
 CC FT DOMAIN 639 645 EXTRACELLULAR (POTENTIAL).
 CC FT REPEAT 148 188 ALPHA-1.
 CC FT REPEAT 520 551 ALPHA-2.
 CC FT DOMAIN 424 430 POLY-GU.
 CC FT CARBOHYD 70 70 N-LINKED (GLCNAC...) (POTENTIAL).
 CC FT CARBOHYD 85 85 N-LINKED (GLCNAC...) (POTENTIAL).
 CC SQ SEQUENCE 645 AA; 71914 MW; C62DEB6CBA4A01C96 CRC64;
 CC -----
 CC Query Match 0.7%; Score 7; DB 1; Length 645;
 CC Best Local Similarity 100.0%; Pred. No. 78;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC QY 790 LGIPDI 796
 CC DB 507 LGIPDI 513
 CC -----
 CC RESULT 40
 CC ID CFPA TREPA STANDARD; PRT; 677 AA.
 CC AC 056336;
 CC DT 15-DEC-1998 (Rel. 37, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Cytoplasmic filament protein A.
 CC GN CFPA OR TP83 OR TP0748.

OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OK NCBI_TaxId=160;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21 AND 146-161.
 RC STRAIN=Nichols;
 RX MEDLINE=9623603; PubMed=8655496;
 RA You Y., Elmore S., Colton L.L., Mackenzie C., Stoops J.K.,
 RA Weinstein G.M., Norris S.J.,
 RT Characterization of the cytoplasmic filament protein gene (cfta) of
 RL Treponema pallidum subsp. pallidum.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=96332770; PubMed=965876;
 RA Frazer C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khakhal H., Richardson D., Howell J.K., Chidambaram M., Uetpack T.,
 RA McDonald L., Ariach P., Bowman C., Colton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete."
 RL Science 281:375-388 (1998).
 CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FILAMENTS THAT RUN THE
 CC LENGTH OF THE ORGANISM JUST UNDERNEATH THE CYTOPLASMIC MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: AN ARRAY OF 4 TO 6 FILAMENTS LIE IN CLOSE
 CC APPosition TO THE INNER MEMBRANE AND ARE ALWAYS LOCALIZED DIRECTLY
 CC UNDERNEATH THE CORRESPONDING GROUP OF PERIPLASMIC FLAGELLA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U32683; AAB0757.1; -;
 CC DR EMBL; A5001246; AAC65715.1; -;
 CC DR TIGR; TP0748;
 CC KW Structural protein; Antigen; Coiled coil; Complete proteome.
 CC FT INIT_MET 0 0
 CC FT DOMAIN 23 48 COILED COIL (POTENTIAL).
 CC FT DOMAIN 289 310 COILED COIL (POTENTIAL).
 CC FT DOMAIN 396 416 COILED COIL (POTENTIAL).
 CC FT DOMAIN 521 539 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 677 AA; 78408 MW; 7F77A14421AB7948 CRC64;
 CC -----
 CC Query Match 0.7%; Score 7; DB 1; Length 677;
 CC Best Local Similarity 100.0%; Pred. No. 81;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC QY 197 DSIIFE 203
 CC DB 264 DSIIFE 270
 CC -----
 CC RESULT 41
 CC ID NAH4 RAT STANDARD; PRT; 717 AA.
 CC AC P26434;
 CC DT 01-AUG-1992 (Rel. 23, Created)
 CC DT 01-AUG-1992 (Rel. 23, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Sodium/hydrogen exchanger 4 (Na(+)/H(+) exchanger 4) (NHE-4).
 CC GN SLC9A4 OR NHE4.
 CC OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxId=10116;

RN SREGION FROM N.A.
 RP STRAIN=Sprague-Dawley; TISSUE=Stomach;
 RC MEDLINE=92250539; PubMed=1577762;
 RX Orlowski J., Kandaany R.A., Shull G.B.;
 RA "Molecular cloning of putative members of the Na/H exchanger gene
 RT family. cDNA cloning, deduced amino acid sequence, and mRNA tissue
 RT expression of the rat Na/H exchanger NHE-1 and two structurally
 RT related proteins.";
 RU J. Biol. Chem. 267:9331-9339(1992).
 CC -1- FUNCTION: INVOLVED IN PH REGULATION TO ELIMINATE ACIDS GENERATED
 CC BY ACTIVE METABOLISM OR TO COUNTER ADVERSE ENVIRONMENTAL
 CC CONDITIONS. MAJOR PROTON EXTRUDING SYSTEM DRIVEN BY THE INWARD
 CC SODIUM ION CHEMICAL GRADIENT. PLAYS AN IMPORTANT ROLE IN SIGNAL
 CC TRANSDUCTION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, FOLLOWED BY COLON
 CC AND SMALL INTESTINE. LESSER AMOUNTS WERE FOUND IN KIDNEY, BRAIN,
 CC UTERUS AND SKELETAL MUSCLE.
 CC -1- PTM: PHOSPHORYLATED (POSSIBLE).
 CC -1- SIMILARITY: BELONGS TO THE NA(+) / H(+) EXCHANGER FAMILY.
 CC -1- CAUTION: THE NUMBER, LOCALIZATION AND DENOMINATION OF HYDROPHOBIC
 CC DOMAINS IN THE NA(+) / H(+) EXCHANGERS VARY AMONG AUTHORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial/
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M85301; AAA41703.1; -.
 DR PIR: C40204; C40204.
 DR InterPro: IPR000676; NaH_Exchange.
 DR InterPro: IPR004709; NaH_exchange3.
 DR Pfam: PF00999; Na_H_Exchange; 1.
 DR PRINTS: PR01084; NAHEXCHNGR.
 DR TRAFRAME: TRG00840; b_cpai; 1.
 DR TransMem: Glycoprotein; Transport; Antiport; Sodium transport;
 KW Multi-gene family; Phosphorylation.
 FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 14 28 A (M1) HYDROPHOBIC.
 FT DOMAIN 29 69 B (M2) HYDROPHOBIC.
 FT DOMAIN 70 90 C (M3) (POTENTIAL).
 FT DOMAIN 91 94 C (M3) (POTENTIAL).
 FT TRANSMEM 115 127 D (M4) (POTENTIAL).
 FT TRANSMEM 128 148 D (M4) (POTENTIAL).
 FT TRANSMEM 149 154 E (M5) (POTENTIAL).
 FT TRANSMEM 155 175 E (M5) (POTENTIAL).
 FT TRANSMEM 176 191 F (M5A) (POTENTIAL).
 FT TRANSMEM 192 216 F (M5B) (POTENTIAL).
 FT TRANSMEM 217 225 G (M5B) (POTENTIAL).
 FT TRANSMEM 226 247 G (M5B) (POTENTIAL).
 FT TRANSMEM 248 269 H (M6) (POTENTIAL).
 FT TRANSMEM 270 290 H (M6) (POTENTIAL).
 FT TRANSMEM 291 304 I (M7) (POTENTIAL).
 FT TRANSMEM 305 325 I (M7) (POTENTIAL).
 FT TRANSMEM 326 356 J (M8) (POTENTIAL).
 FT TRANSMEM 357 377 J (M8) (POTENTIAL).
 FT TRANSMEM 378 384 K (M9) (POTENTIAL).
 FT TRANSMEM 385 405 K (M9) (POTENTIAL).
 FT TRANSMEM 406 420 L, HYDROPHOBIC.
 FT TRANSMEM 421 441 L, HYDROPHOBIC.
 FT TRANSMEM 442 450 M (M10) (POTENTIAL).
 FT TRANSMEM 451 471 M (M10) (POTENTIAL).
 FT TRANSMEM 472 717 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 342 342 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 717 AA; 81522 MW; 4EBBEC7D782753 CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 85;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 147 AAE8TGN 153
 DB 675 AAE8TGN 681
 RESULT 42
 ID PRIA_ECOLI STANDARD; PRT; 732 AA.
 AC P17888;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Primosomal protein N' (Replication factor Y).
 GN PRIA OR B3935.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.
 RX MEDLINE=90280427; PubMed=2162050.
 RA Lee E.H., Masai H., Allen G.C., Jr., Kornberg A.;
 RT "The priA gene encoding the primosomal replicative n' protein of
 RT Escherichia coli.";
 RU Proc. Natl. Acad. Sci. U.S.A. 87:4620-4624(1990).
 RN (2)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 3-16.
 RX MEDLINE=90280426; PubMed=2162049;
 RA Nure P., Digate R., Zavitz K., Mariani K.;
 RT "Molecular cloning and DNA sequence analysis of Escherichia coli
 RT priA, the gene encoding the primosomal protein replication factor
 RT Y.";
 RU Proc. Natl. Acad. Sci. U.S.A. 87:4615-4619(1990).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12 / W61655;
 RX MEDLINE=93347969; PubMed=8346018;
 RA Plunkett G. III, Burdland V.D., Daniels D.L., Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
 RT region from 87.2 to 89.2 minutes.";
 RU Nucleic Acids Res. 21:3391-3398(1993).
 RN (4)
 RP SIMILARITY TO DEAD-BOX HELICASES.
 RX MEDLINE=92107690; PubMed=1662369;
 RA Oucunias C.A., Blencowe B.U.;
 RT "Bacterial DNA replication initiation factor priA is related to
 RT proteins belonging to the 'DEAD-box' family.";
 RU Nucleic Acids Res. 19:6953-6953(1991).
 CC -1- FUNCTION: RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX SSDNA;
 CC THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS PRIA AND
 CC PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE SUBSEQUENT
 CC ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN FUNCTIONS AS A
 CC HELICASE WITHIN THE PRIMOSOME.
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. PRIA SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial/
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M3293; AAA24416.1; -.
 DR EMBL: D00616; BAB00491.1; -.
 DR EMBL: L19201; AAB03067.1; -.
 DR EMBL: A8000467; AAC76917.1; -.
 DR PIR: A35505; A35505.
 DR PIR: A35506; A35506.
 DR ECODBASE: H096.8; 6TH EDITION.
 DR Ecogene; Egi0763; priA.

```
EMBL; AE009762; AAL62720.1; --
InterPro; IPR004543; AEF-2
InterPro; IPR000640; EFG_C
InterPro; IPR004161; BFTU_D2
InterPro; IPR000785; EFG_GPBind
InterPro; IPR005225; Small_Grp
Pfam; PF00679; EFG_C; 1
Pfam; PF00009; BFTU_EFTU; 1
Pfam; PF03144; Grp_EFTU_D2; 1.
```

[illegible]

```

FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 597 597 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 777 AA, 89127 MW, 9B328C3A24A1E7E6 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 777;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 362 EDLEGAV 368
Db 242 EDLEGAV 248

RESULT 45
RA16_SCHPO STANDARD; PRT; 892 AA.
ID RA16_SCHPO
AC P36617.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA repair protein rad16.
GN RAD16 OR RAD10 OR RAD20 OR SW19 OR SPCC970.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCB1_TaxID=4896;
[1]
RP MEDLINE=94158677; PubMed=8114734;
RX Carr A.M., Schmidt H., Kirchhoff S., Muriel W.J., Sheldrick K.S.,
RA Griffiths D.J., Basmacioglu C.N., Subramani S., Clegg M., Nasim A.,
RA Lehmann A.R.;
RT "The rad16 gene of Schizosaccharomyces pombe: a homolog of the RAD1
RT gene of Saccharomyces cerevisiae."
RL Mol. Cell. Biol. 14:2029-2040(1994).
RN
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby I., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moutle S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voickert G., Aert R., Robben J., Grynopre B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,
RA Gabel C., Fuchs W., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado U., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forbush S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpivovskii G.V., Useery D., Barrett B.G., Nure P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- FUNCTION: INVOLVED IN NUCLEOTIDE EXCISION REPAIR OF DNA DAMAGED
CC WITH UV LIGHT, BUZZY ADDUCTS, OR CROSS-LINKING AGENTS.
CC ENDONUCLEASE THAT SPECIFICALLY DEGRADATES SINGLE-STRANDED DNA.
CC -1- SUBUNIT: HETERODIMER COMPOSED OF RAD16 AND SW10.

```

```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE XPF/ERCC4/RAD1/RAD16 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, X71595; CA50599.1; -.
CC EMBL, AL031530; CA20694.1; -.
CC InterPro: IPR004016; ERCC4.
CC Pfam: PF02732; ERCC4; 1.
CC TIGRFAMs: TIGR00596; rad1; 1.
CC DNA repair; DNA-binding; Nuclear protein; Hydrolase; Nuclease;
CC Endonuclease; Phosphorylation.
CC MOD_RSS 71 71 PHOSPHORYLATION (BY CK2) (POTENTIAL).
CC VARIANT 674 674 E -> K (IN SW19 ALLELE).
CC SEQUENCE 892 AA, 102073 MW, 6C0F4B51475333AC CRC64;

Oy 826 VPGSILL 832
Db 47 VPGSILL 53

Query Match 0.7%; Score 7; DB 1; Length 892;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 46
YH6_YEAST STANDARD; PRT; 944 AA.
ID YH6_YEAST
AC Q03631.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative 107.6 kDa transcriptional regulatory protein in CPR3-HMG1
DE intergenic region.
GN YH076C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCB1_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972.
RC Brown D., Bowman S., Barrett B.G., Rajandream M.A.;
RT Submitted (OCT-1994) to the EMBL/Genbank/DBJ databases.
RL -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, Z46373; CA86502.1; -.
CC SGD; S0004541; YML076C.
CC InterPro: IPR001138; Fung1_TYR.
CC SMART; SM00066; GAL4; 1.
CC PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
CC PROSITE; PS00463; ZN2_CY6_FUNGAL_2; FALSE NEG.
CC Hypothetical protein; Transcription regulation; DNA-binding; Zinc;
CC Nuclear protein; Metal-binding.
CC DNA BIND 76 109 ZN(2)-CYS(6), FUNGAL-TYPE.
CC SEQUENCE 944 AA, 107560 MW, 020A56745DF52CCC CRC64;

```

Query Match 0.7%; Score 7; DB 1; Length 944;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 264 IAITGVA 270
Db 5 IAITGVA 11

RESULT 47

POP1_HUMAN

ID POP1_HUMAN STANDARD; PRT; 1024 AA.

AC Q95757; Q15037;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Ribonucleases P/MRP protein subunit POP1 (EC 3.1.26.5) (hPOP1).

GN POP1 OR KIA0061.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI_TaxID=9606;

[1]

RP SEQUENCE OF 1-126 FROM N.A.

RX MEDLINE=97076154; PubMed=8918471;

RA Lygerou Z., Pluk H., Van Ventrpool J. W.J., Seraphin B.;

RT "hPOP1: an autoantigenic protein subunit shared by the human RNase P

and RNase MRP ribonucleoproteins.";

RL EMBO J. 15:5936-5948(1996).

[2]

RP SEQUENCE OF 122-1024 FROM N.A.

RX TISSUE-Bone marrow;

RC MEDLINE=96051398; PubMed=7584044;

RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,

RT Seki N., Kawabayashi Y., Ishikawa K.-I., Tabata S.;

RT "Prediction of the coding sequences of unidentified human genes. II.

The coding sequences of 40 new genes (KIA00041-KIA00080) deduced by

analysis of cDNA clones from human cell line KG-1.";

RL DNA Res. 1:223-229(1994).

CC -1- FUNCTION: COMPONENT OF RIBONUCLEASE P, A PROTEIN COMPLEX THAT

GENERATES MATURE TRNA MOLECULES BY CLEAVING THEIR 5' ENDS. ALSO A

COMPONENT OF RNASE MRP.

CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-

extra-nucleotide from trna precursor.

CC -1- SUBUNIT: RNASE P CONSISTS OF A RNA MOIETY AND AT LEAST 8 PROTEIN

SUBUNITS, POP1, RPP14, RPP20/POP7, RPP29/POP4, RPP30,

RPP38 AND RPP40.

CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X99302; CAA67684.1; -

DR EMBL; D31765; BAA06543.1; -

DR MIM; 602486; -

KW Hydrolyase; Nuclear protein; trna processing.

SQ SEQUENCE 1024 AA; 114708 MW; A1DB872F3B940002 CRC64;

Db 492 GITSPE 783

OY 777 GITSPE 783

RESULT 48

Query Match 0.7%; Score 7; DB 1; Length 1040;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 RLMRLLL 30
Db 565 RLMRLLL 571

RESULT 49

POL1_EEYV8

ID POL1_EEYV8 STANDARD; PRT; 1254 AA.

AC P05674;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Structural polypeptide (P130) [contains: Coat protein C (EC 3.4.21.-)

DE (Capid protein C); Spike glycoprotein E3; Spike glycoprotein E2;

DE 6 kDa peptide; Spike glycoprotein E1].

OS Venezuelan equine encephalitis virus (strain TC-83).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;

NCBI_TaxID=11037;

[1]

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RX STRAIN=Sprague-Dawley; TISSUE=Liver;

RA MEDLINE=91009139; PubMed=2211613;

RT Bischoff J., Moremen K., Lodish H.F.;

RT "Isolation, characterization, and expression of cDNA encoding a rat

liver endoplasmic reticulum alpha-mannosidase.";

RL J. Biol. Chem. 265:17110-17117(1990).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-

mannose residues in alpha-D-mannosides.

CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum.

CC -1- PFM: A SOLUBLE ISOFORM OF THE RAT LIVER MANNOSEDASE IS SUPPOSED TO

BE DERIVED FROM ENDOPLASMIC RETICULUM MEMBRANE ALPHA-MANNOSEDASE

BY PROTEOLYSIS.

CC -1- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M57547; AAA41565.1; -

DR PIR; A38306; A38306.

DR InterPro; IPR000602; Glyco hydro 38.

DR Pfam; PF01074; Glyco hydro 38; 1

KW Hydrolyase; Glycosidase; Endoplasmic reticulum.

SQ SEQUENCE 1040 AA; 115970 MW; 144F6EB97C8F7A7 CRC64;

Db 565 RLMRLLL 571

OY 24 RLMRLLL 30

Query Match 0.7%; Score 7; DB 1; Length 1040;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 565 RLMRLLL 571

OY 24 RLMRLLL 30

RESULT 49

POL1_EEYV8

ID POL1_EEYV8

AC P05674;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Structural polypeptide (P130) [contains: Coat protein C (EC 3.4.21.-)

DE (Capid protein C); Spike glycoprotein E3; Spike glycoprotein E2;

DE 6 kDa peptide; Spike glycoprotein E1].

OS Venezuelan equine encephalitis virus (strain TC-83).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;

NCBI_TaxID=11037;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86306669; PubMed=3755750;

RA Johnson B.J.B., Kinney R.W., Kost C.L., Trent D.W.;

RT "Molecular determinants of alphavirus neurovirulence: nucleotide and

deduced protein sequence changes during attenuation of Venezuelan

equine encephalitis virus.";

```

RL J. Gen. Virol. 67:1951-1960(1986).
CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC HEMAGGLUTININ.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04368; CAA27883.1; -.
CC PIR: A27871; VHWVVE.
CC HSSP: P03315; 1VCP.
CC MEROPS: S03.001; -.
CC InterPro: IPR002548; Alpha_E1_glycop.
CC InterPro: IPR000936; Alpha_E2_glycop.
CC InterPro: IPR002533; Alpha_E3_glycop.
CC InterPro: IPR001836; Alpha_core.
CC InterPro: IPR000930; Togavirin.
CC Pfam: PF00943; Alpha_E2_glycop; 1.
CC Pfam: PF00944; Alpha_core; 1.
CC Pfam: PF01563; Alpha_E3_glycop; 1.
CC Pfam: PF01589; Alpha_E1_glycop; 1.
CC DR PRINTS: PR00798; TOGAVIRIN.
CC KW Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
CC Serine protease.
CC FT CHAIN 1 275 COAT PROTEIN C.
CC FT CHAIN 276 334 SPIKE GLYCOPROTEIN E3.
CC FT CHAIN 335 757 SPIKE GLYCOPROTEIN E2.
CC FT CHAIN 758 812 6 KDA PEPTIDE.
CC FT CHAIN 813 1254 SPIKE GLYCOPROTEIN E1.
CC FT ACT_SITE 152 152 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 158 158 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT TRANSMEM 702 718 POTENTIAL.
CC FT TRANSMEM 774 790 POTENTIAL.
CC FT TRANSMEM 792 808 POTENTIAL.
CC FT TRANSMEM 1231 1247 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 1254 AA; 138485 MW; 7615698519A529F6 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 1254;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 224 TEKMEF 230
Db 656 TEKMEF 662

RESULT 50
POLs_EEYVM STANDARD; PRT; 1254 AA.
ID POLS_EEYVM Q66587; Q66588; Q66589; Q66590; Q66591;
AC P36331; Q66587; Q66588; Q66589; Q66590; Q66591;
DT 01-JUN-1994 (Rel. 29, last sequence update)
DT 01-JUN-1994 (Rel. 29, last sequence update)
DE 16-OCT-2001 (Rel. 40, last annotation update)
DE Structural polyprotein (P130) [contains: Coat protein C (EC 3.4.21.-)
DE (capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
DE 6 kDa peptide; Spike glycoprotein E1].
OS Venezuelan equine encephalitis virus (strain Mena II).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OC NCBI_TaxID=36384;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93187617; PubMed=8445371;
RA Sneider J.M., Kinney R.M., Tsuchiya K.R., Trent D.W.;
RT "Molecular evidence that epizootic Venezuelan equine encephalitis
RT (VEE) I-AB viruses are not evolutionary derivatives of enzootic VEE
RT subtype I-E or II viruses."
RL J. Gen. Virol. 74:519-523(1993).
CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC HEMAGGLUTININ.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04599; AAA42980.1; -.
CC EMBL: L04599; AAA42989.1; ALT_TERM.
CC EMBL: L04599; AAA42991.1; ALT_SEQ.
CC EMBL: L04599; AAA42992.1; ALT_SEQ.
CC EMBL: L04599; AAA42993.1; ALT_SEQ.
CC PIR: J01979; J01979.
CC HSSP: P03315; 1VCP.
CC MEROPS: S03.001; -.
CC InterPro: IPR002548; Alpha_E1_glycop.
CC InterPro: IPR000936; Alpha_E2_glycop.
CC InterPro: IPR002533; Alpha_E3_glycop.
CC InterPro: IPR001836; Alpha_core.
CC InterPro: IPR000930; Togavirin.
CC Pfam: PF00943; Alpha_E2_glycop; 1.
CC Pfam: PF00944; Alpha_core; 1.
CC Pfam: PF01563; Alpha_E3_glycop; 1.
CC Pfam: PF01589; Alpha_E1_glycop; 1.
CC DR PRINTS: PR00798; TOGAVIRIN.
CC KW Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
CC Serine protease.
CC FT CHAIN 1 274 COAT PROTEIN C.
CC FT CHAIN 275 333 SPIKE GLYCOPROTEIN E3.
CC FT CHAIN 334 756 SPIKE GLYCOPROTEIN E2.
CC FT CHAIN 757 812 6 KDA PEPTIDE.
CC FT CHAIN 813 1254 SPIKE GLYCOPROTEIN E1.
CC FT ACT_SITE 151 151 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 157 157 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 225 225 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT TRANSMEM 701 721 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT TRANSMEM 794 813 POTENTIAL.
CC FT TRANSMEM 1231 1248 POTENTIAL.
CC FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 1254 AA; 138343 MW; FB9DB88265F20211 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 1254;
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 224 TEKMEF 230
Db 655 TEKMEF 661

RESULT 51
POLs_EEYVT
NCBI_TaxID=36384;

```

ID POLS BEVVT STANDARD; PRT; 1254 AA.
 AC P09532; Q88691; Q88692; Q88693; Q88694; Q88695;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polypeptide (P130) [Contains: Coat protein C (EC 3.4.21.-)
 (capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
 6 kDa peptide; Spike glycoprotein E1].
 OS Venezuelan equine encephalitis virus (strain Trinidad donkey).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NC NCBI_TaxID=11038;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86263392; PubMed=3088830;
 RA Kinney R.M., Johnson B.J.B., Brown V.L., Trent D.W.;
 RT "Nucleotide sequence of the 26 S mRNA of the virulent Trinidad donkey
 strain of Venezuelan equine encephalitis virus and deduced sequence
 of the encoded structural proteins.";
 RL Virology 152:400-413 (1986).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89243175; PubMed=2524126;
 RA Kinney R.M., Johnson B.J.B., Welch J.B., Tsuchiya K.R., Trent D.W.;
 RT "The full-length nucleotide sequences of the virulent Trinidad donkey
 strain of Venezuelan equine encephalitis virus and its attenuated
 vaccine derivative, strain TC-83.";
 RL Virology 170:19-30 (1989).
 CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
 CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
 CC HEMAGGLUTININ.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC -----
 DR EMBL; M14937; AAA42997.1; -
 DR EMBL; J04332; AAB02519.1; -
 DR PIR; B31467; VHWVVT.
 DR HSSP; P03315; 1VCP.
 DR MEROPS; S03.001; -
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR000936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR001836; Alpha_core.
 DR InterPro; IPR000930; Togavirin.
 DR Pfam; PF00944; Alpha_E2_glycop; 1.
 DR Pfam; PF00943; Alpha_core; 1.
 DR Pfam; PF01563; Alpha_E3_glycop; 1.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 DR PRINTS; PR00798; TOGAVIRIN.
 DR Coat protein; Polypeptide; Transmembrane; Glycoprotein; Hydrolase;
 KW Serine protease.
 FT CHAIN 1 275
 FT CHAIN 276 334
 FT CHAIN 335 357
 FT CHAIN 358 812
 FT CHAIN 813 1254
 FT ACT_SITE 152 152
 FT ACT_SITE 158 158
 FT ACT_SITE 226 226
 FT TRANSMEM 702 718
 FT TRANSMEM 774 790
 FT TRANSMEM 792 808
 FT TRANSMEM 1231 1247
 FT CARBOHYD 286 286
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 811 811 P -> G (IN REF. 2).
 SQ SEQUENCE 1254 AA; 138390 MW; 1C0F62BE1909664B CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 1254;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 224 TEKMEF 230
 DB 656 TEKMEF 662
 RESULT 52
 POLS BEVVT STANDARD; PRT; 1255 AA.
 ID POLS BEVVT STANDARD; PRT; 1255 AA.
 AC P36329;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polypeptide (P130) [Contains: Coat protein C (EC 3.4.21.-)
 (capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
 6 kDa peptide; Spike glycoprotein E1].
 OS Venezuelan equine encephalitis virus (strain 3880).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NC NCBI_TaxID=36382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93079859; PubMed=1448915;
 RA Kinney R.M., Tsuchiya K.R., Snelder J.M., Trent D.W.;
 RT "Genetic evidence that epizootic Venezuelan equine encephalitis (VEE)
 RT viruses may have evolved from enzootic VEE subtype I-D virus.";
 RL Virology 191:569-580 (1992).
 CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
 CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
 CC HEMAGGLUTININ.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC -----
 DR EMBL; L00930; AAC19325.1; -
 DR PIR; D44213; D44213.
 DR HSSP; P03315; 1VCP.
 DR MEROPS; S03.001; -
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR000936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR001836; Alpha_core.
 DR InterPro; IPR000930; Togavirin.
 DR Pfam; PF00944; Alpha_E2_glycop; 1.
 DR Pfam; PF00943; Alpha_core; 1.
 DR Pfam; PF01563; Alpha_E3_glycop; 1.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 DR PRINTS; PR00798; TOGAVIRIN.
 DR Coat protein; Polypeptide; Transmembrane; Glycoprotein; Hydrolase;
 KW Serine protease.
 FT CHAIN 1 275
 FT CHAIN 276 334
 FT CHAIN 335 357
 FT CHAIN 358 813
 FT CHAIN 814 1255
 FT ACT_SITE 152 152
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT ACT_SITE 158 158 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT TRANSMEM 702 722 POTENTIAL.
 FT TRANSMEM 795 814 POTENTIAL.
 FT TRANSMEM 1232 1249 POTENTIAL.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1255 AA; 138297 MW; 72730E17CABCA310 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 1255;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 224 TEKMEF 230
 DB 656 TEKMEF 662

RESULT 53

POLS_EEVVP STANDARD; PRT; 1255 AA.

ID POLS_EEVVP
 AC P36332;
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 40, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polypeptide (P130) [contains: Coat protein C (EC 3.4.21.-)
 (Capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
 6 kDa peptide; Spike glycoprotein E1].
 DE Venezuelan equine encephalitis virus (strain P676).
 OS Virus; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NC NCB1_TaxID=36385;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93079859; PubMed=1448915;
 RA Kinney R.M., Tsuchiya K.R., Snider J.M., Trent D.W.;
 RT "Genetic evidence that epizootic Venezuelan equine encephalitis (VEE)
 RT viruses may have evolved from enzootic VEE subtype 1-D virus."
 RL Virology 191:569-580(1992).
 CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
 CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
 CC HEMAGGLUTININ.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L04653; AAC19319.1; -
 DR PIR: B44213; B44213.
 DR HSSP: P03315; IVCP.
 DR MEROPS: S03.001; -
 DR InterPro: IPR002548; Alpha_E1_glycop.
 DR InterPro: IPR000936; Alpha_E2_glycop.
 DR InterPro: IPR002533; Alpha_E3_glycop.
 DR InterPro: IPR001836; Alpha_core.
 DR InterPro: IPR000930; Togavirin.
 DR Pfam: PF00943; Alpha_E2_glycop; 1.
 DR Pfam: PF00944; Alpha_E3_glycop; 1.
 DR Pfam: PF01563; Alpha_E3_glycop; 1.
 DR Pfam: PF01569; Alpha_E1_glycop; 1.
 DR PRINTS: PR00798; TOGAVIRIN.
 DR CoaT protein; Polypeptide; Transmembrane; Glycoprotein; Hydrolase;
 KW Serine protease.
 FT CHAIN 1 275 COAT PROTEIN C.

FT CHAIN 276 334 SPIKE GLYCOPROTEIN E3.
 FT CHAIN 335 757 SPIKE GLYCOPROTEIN E2.
 FT CHAIN 758 813 6 KDA PEPTIDE.
 FT CHAIN 814 1255 SPIKE GLYCOPROTEIN E1.
 FT ACT_SITE 152 152 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 158 158 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT TRANSMEM 702 722 POTENTIAL.
 FT TRANSMEM 795 814 POTENTIAL.
 FT TRANSMEM 1232 1249 POTENTIAL.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1255 AA; 138213 MW; 33CD302F5CABE646 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 1255;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 224 TEKMEF 230
 DB 656 TEKMEF 662

RESULT 54

POLS_SINDO STANDARD; PRT; 2514 AA.

ID POLS_SINDO
 AC P27283;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nonstructural polypeptide (P270) [contains: Nonstructural protein
 DE NSP1; Nonstructural NSP2; Nonstructural NSP3; Nonstructural NSP4;
 DE Sindbis virus (subtype Ockelbo / strain Bsdyn 82-5).
 OS Sindbis virus; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NC NCB1_TaxID=31699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91229725; PubMed=1673813;
 RA Shirako Y., Niklasson B., Dalrymple J.M., Strauss E.G., Strauss J.H.;
 RT "Structure of the Ockelbo virus genome and its relationship to other
 RT Sindbis viruses."
 RL Virology 182:753-764(1991).
 CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS
 CC BETWEEN THE CODONS FOR 1898-TYR AND 1899-LEU.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M69205; AAA6972.1; ALT_SEQ.
 DR PIR: A39991; MNWY82.
 DR MEROPS: C09.001; -
 DR InterPro: IPR002589; Alpp.
 DR InterPro: IPR002620; Peptidase_C9.
 DR InterPro: IPR001788; RNA_dep_RNAPol2.
 DR InterPro: IPR002588; V_methyltransf.
 DR InterPro: IPR000606; Viral_helicase1.
 DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransf; 1.
 DR Pfam: PF01661; Alpp; 1.
 DR Pfam: PF01707; Peptidase_C9; 1.
 DR SMART: SM00506; Alpp; 1.
 DR Polypeptide; Nonstructural protein; RNA-binding; Helicase.

```

FT CHAIN 1 540 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 541 1347 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1348 1898 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1899 2514 NONSTRUCTURAL PROTEIN NSP4.
SQ SEQUENCE 2514 AA; 279642 MW; 2F38CE32ACF5EDD CRC64;

Query Match
Best Local Similarity 100.0%; DB 1; Length 2514;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 532 KGRQSYT 538
DB 268 KGRQSYT 274

RESULT 55
FRAP_HUMAN STANDARD; PRT; 2549 AA.
AC P42345; Q9Y413;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE FKBP-rapamycin associated protein (FRAP) (Rapamycin target protein).
GN FRAP1 OR FRAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94277209; PubMed=8008069;
RA Brown E.J., Albers M.W., Shin T.B., Ichikawa K., Keith C.T.,
RA Lane W.S., Schreiber S.L.;
RT "A mammalian protein targeted by G1-arresting rapamycin-receptor
RT complex."
RL Nature 369:756-758 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98317532; PubMed=9653645;
RA Onyango P., Lubysova B., Gardellin P., Kurzbauer R., Welch A.;
RT "Molecular cloning and expression analysis of five novel genes in
RT chromosome 1936."
RL Genomics 50:187-198 (1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 2018-2112.
RX MEDLINE=96279639; PubMed=8662507;
RA Choi J., Chen J., Schreiber S.L., Clardy J.;
RT "Structure of the FKBP12-rapamycin complex interacting with the
RT binding domain of human FRAP."
RL Science 273:239-242 (1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 2018-2112.
RX MEDLINE=99190960; PubMed=10089303;
RA Liang J., Choi J., Clardy J.;
RT "Refined structure of the FKBP12-rapamycin-FRB ternary complex at 2.2
RT A resolution."
RL Acta Crystallogr. D 55:736-744 (1999).
RN [5]
RP FUNCTION: ACTS AS THE TARGET FOR THE CELL-CYCLE ARREST AND
RP IMMUNOSUPPRESSIVE EFFECTS OF THE FKBP12-RAPAMYCIN COMPLEX.
RP SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
RP SIMILARITY: CONTAINS 8 HEAT REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC EMBL: L34075; AAA58486.1; -
CC EMBL: U88966; AAC39933.1; -

```

```

DR PDB; 1FAP; 23-JUL-97.
DR PDB; 1NSG; 18-MAR-98.
DR PDB; 1NOB; 18-NOV-98.
DR GENE; HGNC:3942; FRAP1.
DR MIM; 601231; -.
DR InterPro; IPR003151; FAT.
DR InterPro; IPR003152; FATC.
DR InterPro; IPR000357; HEAT_repeat.
DR InterPro; IPR000403; PI3_F14_Kinase.
DR Pfam; PF00454; PI3_F14_Kinase; 1.
DR Pfam; PF02259; FAT; 1.
DR SMART; SM00146; PI3Kc; 1.
DR PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR PROSITE; PS50290; PI3_4_KINASE_3; 1.
DR PROSITE; PS50077; HEAT_REPEAT; FALSE NEG.
KM Transferase; Kinase; Repeat; 3D-structure.
FT REPEAT 16
FT REPEAT 650 688
FT REPEAT 859 897
FT REPEAT 988 1025
FT REPEAT 1069 1106
FT REPEAT 1109 1148
FT REPEAT 1150 1186
FT DOMAIN 1382 1982
FT DOMAIN 1933 1970
FT DOMAIN 2182 2549
FT CONFLICT 353
FT CONFLICT 359
FT CONFLICT 364
FT CONFLICT 390
FT CONFLICT 430
FT CONFLICT 455
FT CONFLICT 461
FT CONFLICT 482
FT CONFLICT 489
FT CONFLICT 513
FT CONFLICT 539
FT CONFLICT 553
FT CONFLICT 956
FT CONFLICT 1075
FT CONFLICT 1075
SQ SEQUENCE 2549 AA; 288888 MW; 7D9AD6E78482AB4 CRC64;

Query Match
Best Local Similarity 100.0%; DB 1; Length 2549;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 594 EASDVGS 600
DB 575 EASDVGS 581

RESULT 56
7LES DROME STANDARD; PRT; 2554 AA.
AC P13368; Q9U5V7; Q9V236; Q9TY10;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sevenless protein (IEC 2.7.1.112).
GN SEV OR HD-265 OR CG18085.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscophora; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=88282538; PubMed=2840202;

```

RA Basler K., Hafen E.;
RT "Control of photoreceptor cell fate by the sevenless protein requires
RT a functional tyrosine kinase domain.";
RL Cell 54:299-311(1988).
RN [2]
RN SEQUENCE FROM N.A.
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN-Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratunga P.G., Scherer S.E., Li P.W., Hoising R.A., Gale R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandel M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazer P.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helf G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Fiancho C., Baldwin D.,
RA Bailey R.M., Bau A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borotova D., Botchan M.R., Bouck J., Cadieu E., Center A., Chandra I.,
RA Burris K.C., Buesam D.A., Butler H., Cadiou E., Davies P.,
RA Cherry J.M., Cawley S., Dahle K., Davemport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodex C., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkhlov G., Milshina N.V., Nobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusserk D.R., Pacheco J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Styliakos R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [4]
RN SEQUENCE OF 2349-2408 FROM N.A.
RP MEDLINE=98401146; PubMed=9711193;
RX Oates A.C., Wollberg P., Achen M.G., Wilks A.F.;
RT "Sampling the genomic pool of protein tyrosine kinase genes using the
RT polymerase chain reaction with genomic DNA.";
RL Biochem. Biophys. Res. Commun. 249:660-667(1998).
RN [5]
RN IDENTIFICATION OF FN-III REPEATS.
RP MEDLINE=90199889; PubMed=2317871;
RX Norton P.A., Hynes R.O., Rees D.J.G.;
RT "Sevenless: seven found?";
RL Cell 61:15-16(1990).
CC -I- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED TO
CC INSTRUCT A CELL TO DIFFERENTIATE INTO A R7 PHOTORECEPTOR. THE
CC LIGAND FOR SEV IS THE BOSS (BRIDE OF SEVENLESS) PROTEIN ON THE
CC SURFACE OF THE NEIGHBORING R8 CELL.
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.

CC	-1	SUBUNIT: MAY FORM A COMPLEX WITH DRK AND SOS.	INSULIN
CC	-1	SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.	
CC	-1	RECEPTOR SUBFAMILY.	
CC	-1	SIMILARITY: CONTRAINS 7 FIBRONECTIN TYPE III-LIKE DOMAINS.	
CC	-1	CAUTION: UNCLEAR WHETHER THE POTENTIAL MEMBRANE SPANNING REGION NEAR THE N-TERMINUS IS PRESENT AS A TRANSMEMBRANE DOMAIN IN THE NATIVE PROTEIN OR SERVES AS A CLEAVED SIGNAL SEQUENCE.	
CC	-1	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce or send an email to license@db-sib.ch).	
CC			
DR	EMBL	J03158; AAA28882.1; -	ALT_INIT.
DR	EMBL	X13666; CA31960.1; -	
DR	EMBL	X13666; CAB5310.1; -	
DR	EMBL	A6003484; AAF4792.1; ALT_INIT.	
DR	PIR	AJ002917; CAA05752.1; -	
DR	PIR	A28912; TYPEFL.	
DR	HSBP	P1162; IFGK.	
DR	FLYBASE	FBgn003366; sev.	
DR	Interpro	IPR00719; Euk_pkinase.	
DR	Interpro	IPR003961; FN_III.	
DR	Interpro	IPR000033; Ldl_receptor_rep.	
DR	Interpro	IPR002011; RTKinaseII.	
DR	Interpro	IPR002290; Ser_thr_pkinase.	
DR	Interpro	IPR01245; Tyr_pkinase.	
DR	Pfam	PF00041; fn3; 6.	
DR	Pfam	PF00069; pkinase; 1.	
DR	PRINTS	PR00109; TYRKINASE.	
DR	Prodom	PD000001; Euk_pkinase; 1.	
DR	SMART	SM00060; FN3; 6.	
DR	SMART	SM0135; LY; 2.	
DR	SMART	SM00220; S_TKC; 1.	
DR	SMART	SM00219; TYRK; 1.	
DR	PROSITE	PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE	PS00109; PROTEIN KINASE TYR; 1.	
DR	PROSITE	PS00239; RECEPTOR TYR KIN III; 1.	
DR	PROSITE	PS00011; PROTEIN KINASE DOM; 1.	
KW	Transferrase; Tyrosine-protein kinase; Receptor; Vision; Transmembrane; Glycoprotein; ATP-binding; Phosphorylation; Repeat.		
KW	EXTRACELLULAR (POTENTIAL).		
KW	POTENTIAL.		
FT	TRANSSEM	2124 2147	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	2148 2554	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	311 431	FIBRONECTIN TYPE-III 2.
FT	DOMAIN	436 528	FIBRONECTIN TYPE-III 3.
FT	DOMAIN	822 921	FIBRONECTIN TYPE-III 4.
FT	DOMAIN	1298 1392	FIBRONECTIN TYPE-III 5.
FT	DOMAIN	1680 1794	FIBRONECTIN TYPE-III 6.
FT	DOMAIN	1797 1897	FIBRONECTIN TYPE-III 7.
FT	DOMAIN	1898 1988	
FT	DOMAIN	2038 2209	POLY-ARG.
FT	DOMAIN	2209 2485	PROTEIN KINASE.
FT	NP_BIND	2215 2223	ATP (BY SIMILARITY).
FT	BLINDING	2242 2242	ATP (BY SIMILARITY).
FT	MOD_RSS	2380 2380	PHOSPHORYLATION (AUTO-)
FT	CARBOHYD	30 30	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	129 129	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	481 481	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	505 505	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	617 617	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	647 647	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	966 966	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	1228 1228	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	1313 1313	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	1353 1353	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	1550 1550	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	1557 1557	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	1639 1639	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	1725 1725	N-LINKED (GLCNAC . .) (POTENTIAL).

FT CARBOHYD 1756 1756 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1804 1804 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1889 1889 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1947 1947 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2073 2073 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2242 2242 K-M: INACTIVATES THE PROTEIN.
 FT CARBOHYD 392 392 V -> M (IN REF. 1).
 FT CARBOHYD 663 663 A -> T (IN REF. 3).
 FT CARBOHYD 1703 1703 N -> H (IN REF. 3).
 FT CARBOHYD 1731 1731 RG -> KE (IN REF. 3).
 FT CARBOHYD 1741 1741 V -> M (IN REF. 3).
 FT CARBOHYD 1823 1823 E -> Q (IN REF. 2).
 FT CARBOHYD 2271 2271 C -> R (IN REF. 1).
 SQ SEQUENCE 2554 AA; 287022 MW; 09E238A0F27684F8 CRC64;

Query Match Best Local Similarity 0.7%; Score 7; DB 1; Length 2554;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 AIVSSCV 872
 DB 2135 AIVSSCV 2141

RESULT 57

ID 7LES_DROVI STANDARD; PRT; 2594 AA.
 AC P20806;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Sevenless protein (EC 2.7.1.112).

OS Drosophila virilis (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7244;

RP SEQUENCE FROM N.A.

RA MEDLINE=90319110; PubMed=2115169;
 Michael W.M., Bowtell D.D.L., Rubin G.M.;

RT "Comparison of the sevenless genes of Drosophila virilis and
 Drosophila melanogaster";

RT Proc. Natl. Acad. Sci. U.S.A. 87:5351-5353(1990).

CC -1- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED TO
 INSTRUCT A CELL TO DIFFERENTIATE INTO A R7 PHOTORECEPTOR. THE
 LIGAND FOR SEV IS THE BOSS (BRIDE OF SEVENLESS) PROTEIN.

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.

CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
 RECEPTOR SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 7 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -1- CAUTION: UNCLEAR WHETHER THE POTENTIAL MEMBRANE SPANNING REGION
 NEAR THE N-TERMINUS IS PRESENT AS A TRANSMEMBRANE DOMAIN IN THE
 NATIVE PROTEIN OR SERVES AS A CLEAVED SIGNAL SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 or send an email to license@ebi.ac.uk).

CC EMBL; M34545; AAA28883.1; -

DR EMBL; M34544; AAA28883.1; JOINED.

DR PIR; A35774; A35774.

DR HSSP; F11362; IFGK.

DR FLYBASE; FBgn0013140; Dv17sev.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR003961; FN_III.

DR InterPro; IPR000033; Ldl_receptor_rep.
 DR InterPro; IPR002011; Rtkkinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00041; fn3; 6.
 DR Pfam; PF00069; pkinase; 1.
 DR Prodom; P000001; Euk_pkinase; 1.
 DR SMART; SM00060; FN3; 5.
 DR SMART; SM00135; LY; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00239; RECEPTOR TYR_KIN_II; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transferase; Tyrosine-protein kinase; Receptor; Vision; Transmembrane;
 Glycoprotein; ATP-binding; Phosphorylation; Repeat.
 FT DOMAIN 1 2139
 FT TRANSMEM 2140 2163
 FT DOMAIN 2164 2594
 FT DOMAIN 9 26
 FT DOMAIN 356 459
 FT DOMAIN 464 555
 FT DOMAIN 835 935
 FT DOMAIN 1328 1421
 FT DOMAIN 1706 1816
 FT DOMAIN 1817 1916
 FT DOMAIN 1917 2007
 FT DOMAIN 2057 2063
 FT DOMAIN 2224 2495
 FT NP_BIND 2230 2238
 FT BINDING 2257 2257
 FT MOD_RES 2391 2391
 FT CARBOHYD 77 77
 FT CARBOHYD 401 401
 FT CARBOHYD 508 508
 FT CARBOHYD 532 532
 FT CARBOHYD 641 641
 FT CARBOHYD 667 667
 FT CARBOHYD 778 778
 FT CARBOHYD 797 797
 FT CARBOHYD 874 874
 FT CARBOHYD 980 980
 FT CARBOHYD 1257 1257
 FT CARBOHYD 1344 1344
 FT CARBOHYD 1382 1382
 FT CARBOHYD 1577 1577
 FT CARBOHYD 1587 1587
 FT CARBOHYD 1655 1655
 FT CARBOHYD 1752 1752
 FT CARBOHYD 1776 1776
 FT CARBOHYD 1824 1824
 FT CARBOHYD 1908 1908
 FT CARBOHYD 1966 1966
 FT CARBOHYD 2088 2088
 SQ SEQUENCE 2594 AA; 289130 MW; 77D8A356B8A0BBD CRC64;

Query Match Best Local Similarity 0.7%; Score 7; DB 1; Length 2594;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 AIVSSCV 872
 DB 2150 AIVSSCV 2156

RESULT 58

ID APC_MOUSE STANDARD; PRT; 2845 AA.
 AC Q61315; Q62044;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein) (MAPC).
 GN APC.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10099;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
 RC STRAIN=CS7BL/6J, and CAST/EI; TISSUE=Brain;
 RX MEDLINE=92263101; PubMed=1350108;
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RT Luongo C., Gould K.A., Dove W.F.,
 RT "Multiple intestinal neoplasia caused by a mutation in the murine
 RT homolog of the APC gene."
 RL Science 256:668-670 (1992).
 RN [2]
 RP ERRATUM.
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RA Luongo C., Gould K.A., Dove W.F.;
 RL Science 256:1114-1114 (1992).
 RN [3]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Dicker F., Lambert S., Reimair A., Balhausen W.G.;
 RT "The murine APC gene: alternative splicing of 5' untranslated
 RT region segments."
 RL Submitted (Oct-1993) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=94061824; PubMed=8242607;
 RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
 RT "APC gene messenger RNA: novel isoforms that lack exon 7."
 RL Cancer Res. 53:5589-5591 (1993).
 CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
 CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
 CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATEININ (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
 CC SIMILARITY).
 CC -1- PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
 CC LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M88127; AAB59632.1; -
 CC EMBL; U02937; AAO3443.1; -
 CC HSSP; 002248; 3BCT.
 CC MGD; MGI:86039; APC.
 CC InterPro; IPR000225; Armadillo.
 CC Pfam; PF00514; Armadillo_seg; 6.
 CC SMART; SM00185; ARM; 5.
 CC PROSITE; PS50176; ARM_REPEAT; 1.
 CC Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;
 CC Coiled coil.
 FT DOMAIN 1 61 COILED COIL (POTENTIAL).
 FT DOMAIN 125 245 COILED COIL (POTENTIAL).
 FT DOMAIN 1 728 LEU-RICH.
 FT REPEAT 451 493 ARM 1.
 FT REPEAT 503 545 ARM 2.
 FT REPEAT 546 589 ARM 3.
 FT REPEAT 590 636 ARM 4.
 FT REPEAT 637 681 ARM 5.
 FT REPEAT 682 723 ARM 6.
 FT REPEAT 724 765 ARM 7.
 FT DOMAIN 739 2834 SER-RICH.

FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1864 1891 HIGHLY CHARGED.
 FT VARSPLIC 243 276 MISSING (IN ISOFORM 2 AND ISOFORM 4).
 FT VARSPLIC 310 410 MISSING (IN ISOFORM 3 AND ISOFORM 4).
 FT VARSPLIC 120 120 MISSING (IN STRAIN CAST/EI).
 FT VARSPLIC 493 493 T -> A (IN STRAIN CAST/EI).
 FT VARSPLIC 797 797 V -> I (IN STRAIN CAST/EI).
 FT VARSPLIC 1330 1330 Y -> F (IN STRAIN CAST/EI).
 FT VARSPLIC 1618 1618 A -> T (IN STRAIN CAST/EI).
 FT VARSPLIC 2294 2294 G -> A (IN STRAIN CAST/EI).
 FT VARSPLIC 2496 2496 H -> Q (IN STRAIN CAST/EI).
 FT VARSPLIC 2523 2523 T -> A (IN STRAIN CAST/EI).
 FT VARSPLIC 2813 2813 T -> S (IN STRAIN CAST/EI).
 SQ SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;
 Query Match 0.7%; Score 7; DB 1; Length 2845;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 317 SEKGSS 323
 Db 2687 SEKGSS 2693
 RESULT 59
 ID SACS_HUMAN STANDARD; PRT; 3829 AA.
 AC Q9NZJ4; O94835;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sacsin.
 GN SACS OR KIA0730.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT ALA-2619.
 RX MEDLINE=20120709; PubMed=10655055;
 RA Engert J.C., Berube P., Mercier J., Dore C., Lepage P., Ge B.,
 RA Bouchard J.-P., Mathieu J., Melancon S.B., Schalling M., Lander E.S.,
 RA Morgan K., Hudson T.U., Richter A.;
 RT "ARSACS, a spastic ataxia common in northeastern Quebec, is caused by
 RT mutations in a new gene encoding an 11.5-kb ORF."
 RL Nat. Genet. 24:120-125 (2000).
 RN [2]
 RP SEQUENCE OF 2826-3829 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286 (1998).
 CC -1- FUNCTION: MAY FUNCTION IN CHAPERONE-MEDIATED PROTEIN FOLDING.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE CENTRAL NERVOUS
 CC SYSTEM. ALSO FOUND IN SKELETAL MUSCLE AND AT LOW LEVELS IN
 CC PANCREAS.
 CC -1- DISEASE: DEFECTS IN SACS ARE THE CAUSE OF AUTOSOMAL RECESSIVE
 CC SPASTIC ATAXIA OF CHARLEVOIX-SAGUENAY (ARSACS OR SACS). ARSACS IS
 CC AN EARLY ONSET NEURODEGENERATIVE DISEASE WITH HIGH PREVALENCE IN
 CC THE CHARLEVOIX-SAGUENAY-LAC-SAINT-JEAN REGION OF QUEBEC. IT IS
 CC CHARACTERIZED BY ABSENT SENSORY-NERVE CONDUCTION, REDUCED MOTOR-
 CC NERVE VELOCITY AND HYPERMYELINATION OF RETINAL-NERVE FIBERS.
 CC -1- SIMILARITY: CONTAINS 10 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF193556; AAFJ1262.1; -
DR EMBL; AB018273; BAA34450.1; -
DR Genew; HGNC:10519; SACS.
DR MIM; 604490; -
DR MIM; 270550; -
DR InterPro; IPR001623; DnaJ_N.
DR PROSITE; PS00636; DnaJ_1; FALSE_NEG.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chaperone; Polymorphism.
FT DOMAIN 356 3643
FT VARIANT 2619 2619 J-DOMAIN.
FT V -> A.
FT /FTID=VAR 010296.
SQ SEQUENCE 3829 AA; 436972 MW; 7ABE90311B1EBE91 CRC64;

Oy 652 SUCYND 658
   |||||
Db 1715 SUCYND 1721

RESULT 60
SACS_MOUSE
ID ID_SACS_MOUSE STANDARD; PRT; 3830 AA.
AC 09JUL98;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sacsin.
GN SACS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RX [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=20120709; PubMed=10655055;
RA Engert J.C., Berube P., Mercier J., Dore C., Lepage P., Ge B.,
RA Bourhard J.-P., Mathieu J., Melancon S.B., Schalling M., Lander E.S.,
RA Morgan K., Hudson T.J., Richter A.;
RT "ARACS, a specific ataxia common in northeastern Quebec, is caused by
RT mutations in a new gene encoding an 11.5-kb ORF.";
RL Nat. Genet. 24:120-125(2000).
CC -I- FUNCTION: MAY FUNCTION IN CHAPERONE-MEDIATED PROTEIN FOLDING.
CC -I- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on items
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF193557; AAFJ1263.1; -
DR MGD; MGI:1354724; Sacs.
DR InterPro; IPR001623; DnaJ_N.
DR PROSITE; PS00636; DnaJ_1; FALSE_NEG.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chaperone.
FT DOMAIN 3557 3644 J-DOMAIN.
FT SEQUENCE 3830 AA; 436750 MW; B3F9DA5A3B0C3EF1 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 3830;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY	652	SLOCYNDC	658
D0	1715	SLOCYNDG	1721

RESULT 61

ID	_DYHC_RAT	STANDARD;	PRT;	4644_AA.
DT	P38650; 063178;			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Dynein heavy chain, cytosolic (DYHC) (Cytoplasmic dynein heavy chain) (MAP 1C).			
GN	DNCH1 OR DNCH1 OR DNEC1 OR MAPLC.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Rattus.			
OX	NCHI_TaxID=10116;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Mistari; TISSUE=Brain;			
RC	MEDLINE=93376715; PubMed=769037;			
RX	Zhang Z., Tanaka Y., Nonaka S., Aizawa H., Kawasaki H., Nakata T.,			
RA	Hirokawa N.;			
RA	"The primary structure of rat brain (cytoplasmic) dynein heavy chain,			
RT	a cytoplasmic motor enzyme.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 90:7928-7932(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Sprague-Dawley; TISSUE=Brain;			
RC	MEDLINE=93264075; PubMed=768432;			
RX	Mikami A., Paschal B.M., Mazumdar M., Vallee R.B.;			
RT	"Molecular cloning of the retrograde transport motor cytoplasmic			
RT	dynein (MAP 1C).";			
RL	Neuron 10:787-796(1993).			
-I-	FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A			
CC	MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND			
CC	ORGANELLES ALONG MICROTUBULES.			
-I-	SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF			
CC	INTERMEDIATE AND LIGHT CHAINS.			
-I-	SUCCELLULAR LOCATION: Cytoplasmic.			
CC	-I- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.			

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

DR	EMBL; D13896; BAAD2996.1; -.			
DR	EMBL; L08505; AAAA1103.1; -.			
DR	PIR; A38905; A38905.			
DR	InterPro; IPR004273; Dynein heavy.			
KM	Pfam; PF03028; Dynein heavy; 1.			
FT	Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.			
FT	DOMAIN 48 69 COILED COIL (POTENTIAL).			
FT	DOMAIN 179 200 COILED COIL (POTENTIAL).			
FT	DOMAIN 453 476 COILED COIL (POTENTIAL).			
FT	DOMAIN 541 564 COILED COIL (POTENTIAL).			
FT	DOMAIN 1169 1201 COILED COIL (POTENTIAL).			
FT	DOMAIN 1229 1250 COILED COIL (POTENTIAL).			
FT	DOMAIN 1355 1371 COILED COIL (POTENTIAL).			
FT	DOMAIN 2012 2040 MICROTUBULE-BINDING (POTENTIAL).			
FT	DOMAIN 3187 3273 COILED COIL (POTENTIAL).			
FT	DOMAIN 3394 3498 COILED COIL (POTENTIAL).			
FT	DOMAIN 3735 3798 COILED COIL (POTENTIAL).			
FT	NP_BIND 1904 1911 ATP (POTENTIAL).			
FT	NP_BIND 2222 2229 ATP (POTENTIAL).			
FT	NP_BIND 2593 2600 ATP (POTENTIAL).			
FT	NP_BIND 2935 2942 ATP (POTENTIAL).			

FT CONFLICT 1024 1025 SR -> MP (IN REF. 2).
 FT CONFLICT 1772 1772 N -> D (IN REF. 2).
 FT CONFLICT 2098 2098 P -> A (IN REF. 2).
 FT CONFLICT 2139 2139 F -> V (IN REF. 2).
 FT CONFLICT 2175 2175 D -> A (IN REF. 2).
 FT CONFLICT 2185 2185 K -> Q (IN REF. 2).
 FT CONFLICT 2366 2366 L -> V (IN REF. 2).
 FT CONFLICT 2382 2382 T -> S (IN REF. 2).
 FT CONFLICT 3219 3219 G -> A (IN REF. 2).
 FT CONFLICT 3219 3219 A -> D (IN REF. 2).
 FT CONFLICT 4131 4131 R -> K (IN REF. 2).
 FT CONFLICT 4366 4366 F -> S (IN REF. 2).
 FT CONFLICT 4511 4511 A -> G (IN REF. 2).
 SQ SEQUENCE 4644 AA; 532240 MM; 8C6ABDBEDF875D82 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 4644;
 Best Local Similarity 100.0%; Pred. No. 4, 4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SARVGR 17
 |||||
 Db 720 SARVGR 726

RESULT 62

PSBL_ARATH STANDARD; PRT; 37 AA.

AC P29301;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Photosystem II reaction center L protein (PSII 5 kDa protein).
 GN PSBL.
 OS Arabidopsis thaliana (Mouse-ear cress), and
 OS Antirrhinum majus (Garden snapdragon).
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702, 4151;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=A.majus; STRAIN=cv. Sippe 50; TISSUE=leaf;
 RA MEDLINE=92191997; PubMed=1547774;
 RA Kuda J., Igloi G.L., Metzlaef M., Hagemann H., Kessel H.;
 RT "RNA editing in tobacco chloroplasts leads to the formation of a
 RT translatable psbl mRNA by a C to U substitution within the initiation
 RT codon.";
 RL EMBL J. 11:1099-1103(1992).

RP SEQUENCE FROM N.A.
 RC SPECIES=A.thaliana; STRAIN=cv. Columbia;
 RX MEDLINE=20039611; PubMed=10574454;
 RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
 RT "Complete structure of the chloroplast genome of Arabidopsis
 RT thaliana.";
 RL DNA Res. 6:283-290(1999).
 CC -1- FUNCTION: NOT KNOWN. POSSIBLY A PERIPHERAL COMPONENT OF CYTOCHROME
 CC B559.

-1- SIMILARITY: BELONGS TO THE PSBL FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X63206; CAA44889.1; -
 DR EMBL; AP000423; BAA84400.1; -
 DR PIR; S19786; P2SKL.
 DR InterPro; IPR003372; PSII_Psbl.

DR Pfam; PF02419; Psbl; 1.
 KW Photosynthesis; Photosystem II; Chloroplast.
 FT INIT MET 0
 SQ SEQUENCE 37 AA; 4339 MM; CC537AEC4C630A84 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 37;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 SVELNR 237
 |||||
 Db 9 SVELNR 14

RESULT 63

TX21_SELHU STANDARD; PRT; 37 AA.

AC P82959;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Huwentoxin-II isoform 1 (HwTx-II).
 OS Selenocosmia huwena (Chinese bird spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Mygalomorphae; Theraphosidae; Ornithoconus.
 ON NCBI_TaxID=29017;
 RX [1]
 RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
 RC TISSUE=Venom;
 RX MEDLINE=99351628; PubMed=10424342;
 RA Shu Q., Liang S.-P.;
 RT "Purification and characterization of huwentoxin-II, a neurotoxic
 RT peptide from the venom of the Chinese bird spider Selenocosmia
 RT huwena.";
 RL J. Pept. Res. 53:486-491(1999).
 CC -1- FUNCTION: LETHAL NEUROTOXIN. BINDS TO THE NICOTINIC ACETYLCHOLINE
 CC RECEPTOR. BLOCKS NEUROMUSCULAR TRANSMISSION. ACTS COOPERATIVELY TO
 CC POTENTIATE THE ACTIVITY OF HUWENTOXIN-I.
 CC -1- SUBUNIT: ISOFORM 1 AND ISOFORM 2 MAY DIMERIZE.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC -1- MASS SPECTROMETRY: MW=4290.3; METHOD=MALDI.
 KW -1- SIMILARITY: BELONGS TO THE HUWENTOXIN-II FAMILY.
 RX Toxin; Neurotoxin; Postsynaptic neurotoxin.
 FT DISULFID 4
 FT DISULFID 8
 FT DISULFID 23
 FT DISULFID 34
 SQ SEQUENCE 37 AA; 4290 MM; E0F9F84321AD6382 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 37;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 ECSFSC 95
 |||||
 Db 3 ECSFSC 8

RESULT 64

TX22_SELHU STANDARD; PRT; 37 AA.

AC P82960;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Huwentoxin-II isoform 2 (HwTx-II).
 OS Selenocosmia huwena (Chinese bird spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Mygalomorphae; Theraphosidae; Ornithoconus.
 ON NCBI_TaxID=29017;
 RX [1]
 RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
 RC TISSUE=Venom;

RX MEDLINE=99351628; PubMed=10424342;
 RA Shu O., Liang S.-P.;
 RT "Purification and characterization of huwentoxin-II, a neurotoxic
 RT peptide from the venom of the Chinese bird spider *Selenocosmia
 huwena*.";
 RL J. Pept. Res. 53:486-491(1999).
 CC -1- FUNCTION: LETHAL NEUROTOXIN. BINDS TO THE NICOTINIC ACETYLCHOLINE
 CC RECEPTOR. BLOCKS NEUROMUSCULAR TRANSMISSION. ACTS COOPERATIVELY TO
 CC POTENTIATE THE ACTIVITY OF HUWENTOXIN-I.
 CC -1- SUBUNIT: ISOFORM 1 AND ISOFORM 2 MAY DIMERIZE.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC -1- MASS SPECTROMETRY: MW=4305.2; METHOD=MALDI.
 CC -1- SIMILARITY: BELONGS TO THE HUWENTOXIN-II FAMILY.
 KW Toxin; Neurotoxin; Postsynaptic neurotoxin.
 FT DISULFID 4 18
 FT DISULFID 8 29
 FT DISULFID 23 34
 SQ SEQUENCE 37 AA; 4305 MW; E0F9F85A0356382 CRC64;
 Query Match 0.6%; Score 6; DB 1; Length 37;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 90 ECFSC 95
 DB 3 ECFSC 8
 RESULT 65
 PSBL_CHLVU
 ID PSBL_CHLVU STANDARD; PRT; 38 AA.
 AC P56339;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Photosystem II reaction center L protein (PSII 5 kDa protein).
 GN PSBL.
 OS Chlorella vulgaris.
 OG Chloroplast.
 OC Eukaryote; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
 OC Chlorellaceae; Chlorella.
 OC NCBI_TaxID=3077;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IAM C-27 / Tamiya;
 RX MEDLINE=97303241; PubMed=9159184;
 RA Matsuura T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
 RA Toudzuki J., Nakashima K., Toudzuki T., Suzuki Y., Hamada A., Ohta T.,
 RA Itamura A., Yoshinaga K., Sugita M.;
 RT "Complete nucleotide sequence of the chloroplast genome from the
 RT green alga *Chlorella vulgaris*: the existence of genes possibly
 RT involved in chloroplast division.";
 RT Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
 RL -1- FUNCTION: NOT KNOWN. POSSIBLY A PERIPHERAL COMPONENT OF CYTOCHROME
 CC B559.
 CC -1- SIMILARITY: BELONGS TO THE PSBL FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB001684; BAA57901.1; -
 DR InterPro: IPR003372; PSII_Psbl.
 DR Pfam; PF02419; Psbl; 1.
 KW Photosynthesis; Photosystem II; Chloroplast.
 SQ SEQUENCE 38 AA; 4388 MW; DF737195D0DEA61D CRC64;
 Query Match 0.6%; Score 6; DB 1; Length 38;

Best Local Similarity 100.0%; Pred. No. 76;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 232 SVELNR 237
 DB 10 SVELNR 15
 RESULT 66
 PSBL_MARPO
 ID PSBL_MARPO STANDARD; PRT; 38 AA.
 AC P12155;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Photosystem II reaction center L protein (PSII 5 kDa protein).
 GN PSBL.
 OS Marchantia polymorpha (Liverwort).
 OG Chloroplast.
 OC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
 OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;
 OC Marchantiaceae; Marchantia.
 OC NCBI_TaxID=3197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89068687; PubMed=3199436;
 RA Fukuzawa H., Kohchi T., Sano T., Shirai H., Umesono K., Inokuchi H.,
 RA Ozeki H., Ohyama K.;
 RT "Structure and organization of *Marchantia polymorpha* chloroplast
 RT genome. III. Gene organization of the large single copy region from
 RT *hbcL* to *trnI* (CAU).";
 RT J. Mol. Biol. 203:333-351(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
 RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
 RA Ozeki H.;
 RT "Chloroplast gene organization deduced from complete sequence of
 RT liverwort *Marchantia polymorpha* chloroplast DNA.";
 RT Nature 322:572-574(1986).
 RL -1- FUNCTION: NOT KNOWN. POSSIBLY A PERIPHERAL COMPONENT OF CYTOCHROME
 CC B559.
 CC -1- SIMILARITY: BELONGS TO THE PSBL FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04465; CAA28099.1; -
 DR PIR; A05048; A05048.
 DR PIR; S01538; S01538.
 DR InterPro: IPR003372; PSII_Psbl.
 DR Pfam; PF02419; Psbl; 1.
 KW Photosynthesis; Photosystem II; Chloroplast.
 SQ SEQUENCE 38 AA; 4479 MW; DF73781163D39D6D CRC64;
 Query Match 0.6%; Score 6; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 232 SVELNR 237
 DB 10 SVELNR 15
 RESULT 67
 PIV6_ADEB2
 ID PIV6_ADEB2 STANDARD; PRT; 43 AA.
 AC Q96627;
 Query Match 0.6%; Score 6; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Minor capsid protein VI precursor (Fragment).
GN PVI
OS Bovine adenovirus type 2 (Mastadenovirus bo2).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
NCBI_Taxid=114429;
RN [1]
RP SEQUENCE FROM N.A.
RA Rusvai M., Harrach B., Banreivi A., Evans P., Benko M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MINOR CAPSID PROTEIN THAT MAY ACT AS A LINK BETWEEN THE
CC EXTERNAL CAPSID AND THE INTERNAL DNA-PROTEIN CORE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U4123; AAB16759.1; -
DR InterPro; IPR004243; MCVI.
DR Pfam; PF02993; MCPVI.
FT PROPEP 1 33 BY SIMILARITY.
FT CHAIN 34 >43 MINOR CAPSID PROTEIN VI.
FT NON TER 43 43
SQ SEQUENCE 43 AA; 4584 MW; 10F7BE9678070306 CRC64;

Query Match
Best Local Similarity 0.6%; Score 6; DB 1; Length 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 NFSALA 677
DB 5 NFSALA 10

RESULT 68
ATP8_YARLI
ID ATP8_YARLI STANDARD; PRT; 48 AA.
AC Q36257;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).
GN ATP8.
OS Yarrowia lipolytica (Candida lipolytica).
OC Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
NCBI_Taxid=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 44601;
RA Matsubara M., Inoue J., Kakehi M., Imanaka T.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; L15359; AAA78260.1; -
DR Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
KW TRANSMEM 13
FT TRANSMEM 32 POTENTIAL.
SQ SEQUENCE 48 AA; 5832 MW; 6076BA02936A0BF7 CRC64;

Query Match
Best Local Similarity 0.6%; Score 6; DB 1; Length 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 GPASLS 138
DB 16 GPASLS 21

RESULT 69
ATP8_PARLI
ID ATP8_PARLI STANDARD; PRT; 54 AA.
AC P12697;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).
GN MTAtp8 OR ATP8.
OS Paracentrotus lividus (Common sea urchin).
OC Mitochondrion.
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinozoa; Euechinozoa; Echinodermata; Echinodermata; Echinodermata;
OC Paracentrotus.
NCBI_Taxid=7656;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8921831; PubMed=2544576;
RA Cantatore P., Roberti M., Rainaldi G., Gadaleta M.N., Saccone C.;
RT "The complete nucleotide sequence, gene organization, and genetic
RT code of the mitochondrial genome of Paracentrotus lividus."
RL J. Biol. Chem. 264:10965-10975(1989).
CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U04815; AAA68136.1; -
DR PIR; F34284; F34284.
DR InterPro; IPR001421; ATPase8_mit.
DR Pfam; PF00895; ATP-synt 8; 1_
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
SQ SEQUENCE 54 AA; 6452 MW; 41E52C808607EA46 CRC64;

Query Match
Best Local Similarity 0.6%; Score 6; DB 1; Length 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 919 TAILLT 924
DB 18 TAILLT 23

RESULT 70
Y737_ARCFU
ID Y737_ARCFU STANDARD; PRT; 67 AA.
AC Q29521;

```

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AF0737.
 GN AF0737.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN
 RP
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RA MEDLINE=98049343; PubMed=9389475;
 RA Klein H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus."
 RL Nature 390:364-370(1997).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AEO01054; AAB90514.1; -
 DR TIGR; AF0737; -
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 67 AA; 7296 MW; 6F103AABBB56630 CRC64;
 SQ
 Query Match 0.6%; Score 6; DB 1; Length 67;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 787 LESLGI 792
 |||||
 Db 28 LESLGI 33

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens S.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Welfjens I., Vanstreets E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabell C., Fuchs M., Fritz C., Holzer E., Moesli D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wandt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
 RA Lucifert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Chuado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC
 CC -1- FUNCTION: PROBABLE COMMON SM PROTEIN, IS FOUND IN U1 AND U2
 CC SNRPS (BY SIMILARITY).
 CC
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC
 CC -1- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL031534; CA119017.1; -
 DR EMBL; AL031534; CAA20721.1; -
 DR InterPro; IPR001163; snRNP_Sm.
 DR Pfam; PF01423; Sm; 1.
 KW Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;
 KW RNA-binding.
 SQ SEQUENCE 78 AA; 8660 MW; 005687FA91F013CA CRC64;
 Query Match 0.6%; Score 6; DB 1; Length 78;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 257 KPVLVR 262
 |||||
 Db 18 KPVLVR 23

RESULT 71
 ID RUXF_SCHPO STANDARD; PRT; 78 AA.
 AC 059734;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable small nuclear ribonucleoprotein F (snRNP-F) (Sm protein F)
 DE (Sm-F) (SMP).
 GN SMF1 OR SPBC3E7.14 OR SPBC4F6.01.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN
 RP
 RC STRAIN=972;
 RA MEDLINE=21846401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RESULT 72
 ID NSGX_HUMAN STANDARD; PRT; 79 AA.
 AC Q9YH64;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Susceptibility protein NSG-X.
 GN Homo sapiens (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN
 RP
 RC TISSUE=Nasopharyngeal carcinoma;
 RA Jun Q., Yang J.B., Li G.Y.;
 RT "A new nasopharyngeal carcinoma associated gene on 9p21-22."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RA Burian D.M., Mitchell N., Roe B.A.;

```

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF211119; AAF22974.1; -
DR EMBL: AC000048; -; NOT_ANNOTATED_CDS.
SQ SEQUENCE 79 AA; 8851 MW; EC94C8BC34FD5F8C CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 RIPRLW 26
Db 64 RIPRLW 69

RESULT 73
YVKA VACCC STANDARD; PRT; 81 AA.
ID YVKA VACCC
AC P20569;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 9.4 kDa protein.
GN K ORF A.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkins M.E., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35027; AAA48007.1; -
DR PIR: G42505; G42505.
KW Hypothetical protein.
SQ SEQUENCE 81 AA; 9386 MW; 42CE3D0531FEA9C4 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 571 KIYSIN 576
Db 73 KIYSIN 78

RESULT 74

```

```

Y67_BPT3
ID Y67_BPT3 STANDARD; PRT; 83 AA.
AC P20330;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Hypothetical gene 6.7 protein.
GN 6.7.
OS Bacteriophage T3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Luria;
RX MEDLINE=90133923; PubMed=2614843;
RA Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RT "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.";
RL J. Mol. Biol. 210:687-701(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17255; CA435150.1; -
DR PIR: S07519; S07519.
KW Hypothetical protein.
SQ SEQUENCE 83 AA; 8848 MW; C4953F5806081581 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 LDDSA 148
Db 64 LDDSA 69

RESULT 75
Y425_TREPA STANDARD; PRT; 86 AA.
ID Y425_TREPA
AC O83440;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0425.
GN TP0425.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98322770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardam J.M., McLeod M.P., Salberg S., Peterson J.,
RA Khailak H., Richardson D.J., Howell J.K., Chidambaram M., Utechtack T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AEO01220; AAC65419.1; -
 DR TIGR; TP0425; -
 KW Hypothetical protein; Complete proteome;
 SQ SEQUENCE 86 AA; 9737 MW; 728CDB709B7132DB CRC64;

Query Match 0.6%; Score 6; DB 1; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 DSTGSR 67
 DB 8 DSTGSR 13

RESULT 76
 ID FXV4_MOUSE STANDARD; PRT; 88 AA.
 AC Q9D2W0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE FXVD domain-containing ion transport regulator 4 precursor (Channel
 DE inducing factor) (ChIF).
 GN FXVD4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SvJ;
 RA Garty H.;
 RT "Genomic sequence of mouse ChIF (FXVD4).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

[2]
 SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schram L.M., Straubli F., Suzuki R., Tomita M., Wagner I., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kanliya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seyer T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontseki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE FXVD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF362729; AAK51508.1; -
 DR EMBL; AK018728; BAB31372.1; -
 DR MGD; MGI:1889005; Fxyd4.
 DR InterPro; IPR000272; ATP1G1_PLM_MAT8.
 DR Pfam; PF02038; ATP1G1_PLM_MAT8; 1.
 DR PROSITE; PS01310; FXVD; 1.
 KW Transmembrane, Signal, Ionic channel, Ion transport.
 FT SIGNAL 1 20
 FT CHAIN 21 88
 FT DOMAIN 21 38
 FT TRANSMEM 39 59
 FT DOMAIN 60 88
 SQ SEQUENCE 88 AA; 9269 MW; 7EB0140941CFE926 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 88;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 ALEASD 597
 DB 17 ALEASD 22

RESULT 77
 ID XHLA_BACSU STANDARD; PRT; 89 AA.
 AC P39798;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein xhla.
 GN XHLA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / S0113;
 RA Krogh S., Joergensen S.T., Diderichsen B., Devine K.M.;
 RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.

[2]
 SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=9505440; PubMed=7921239;
 RA Longchamp P.F., Mauei C., Karamata D.;
 RT "lytic enzymes associated with defective prophages of Bacillus
 RT subtilis: sequencing and characterization of the region comprising
 RT the N-acetylmuramoyl-L-alanine amidase gene of prophage PBSX";
 RL Microbiology 140:1855-1867(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98215178; PubMed=9555893;
 RA Krogh S., Joergensen S.T., Devine K.M.;
 RT "lytic genes of the Bacillus subtilis defective prophage PBSX";
 RL J. Bacteriol. 180:2110-2117(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berrero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Ertter K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizi A., Galleron N.,
 RA Gilm S.Y., Glaeser P., Gottreau A., Goughly E.J., Grand G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kaehara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsleip G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetle D., Potwilk S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vanlier F., Vassarotti A.,
 RA Viali A., Wambutt R., Wedler H., Wedler H., Welteregger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumein E., Yoshikawa H., Zarchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*,"
 RL Nature 390:249-256(1997).
 CC -1- FUNCTION: ASSOCIATED WITH CELL LYSIS UPON INDUCTION OF PBSX.
 CC -1- SIMILARITY: STRONG, TO B.LICHENIFORMIS XPAFI AND XPAFI.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; Z36941; CAA85401.1; -
 CC DR EMBL; L25924; AAA22643.1; -
 CC DR EMBL; Z70177; CAA34047.1; -
 CC DR EMBL; Z29110; CAA3136.1; -
 CC DR PIR; S47314; S47314.
 CC DR Subtilisin; BG10960; xhla.
 CC KW Transmembrane; Complete proteome.
 CC FT TRANSMEM 63 POTENTIAL.
 CC SQ SEQUENCE 89 AA; 9997 MW; EBBB53B7A3B81AEA CRC64;
 CC -----
 CC Query Match 0.6%; Score 6; DB 1; Length 89;
 CC Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC Oy 541 IEENTT 546
 CC DB 52 IEENTT 57
 CC -----
 CC RESULT 78
 CC YALI_BACLI STANDARD; PRT; 89 AA.
 CC AC 099164;
 CC DT 01-OCT-1994 (Rel. 30, Created)
 CC DT 01-OCT-1994 (Rel. 30, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Hypothetical 10.1 kDa protein in ORF3 5'region.
 CC GN XPAFI.
 CC OS *Bacillus licheniformis*.
 CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CC OX NCBI_TaxID=1402;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=MC14;
 CC RX MEDLINE=91237328; PubMed=2033382;
 CC RA Lee J.W.K., Edwards C.W., Hulett F.M.;
 CC RT "Identification of four unique clones encoding 10 kDa proteins from
 CC RT *Bacillus* that cause phenotypic complementation of a *proA* mutant
 CC RT strain of *Bacillus coli*,"
 CC RT J. Gen. Microbiol. 137:667-677(1991).
 CC CC -1- SIMILARITY: TO B.LICHENIFORMIS XPAFI AND TO B.SUBTILIS XHIA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M63942; AAA22886.1; -
 CC DR PIR; B49754; B49754.
 CC KW Hypothetical protein.
 CC SQ SEQUENCE 89 AA; 10071 MW; 6686426FB9356ABE CRC64;
 CC -----
 CC Query Match 0.6%; Score 6; DB 1; Length 89;
 CC Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC Oy 541 IEENTT 546
 CC DB 52 IEENTT 57
 CC -----
 CC RESULT 79
 CC ES6B_MYCTU STANDARD; PRT; 95 AA.
 CC AC 005454;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Putative ESAT-6 like protein 11.
 CC GN RV3890C OR MT4005 OR MTCY15F10.22.
 CC OS *Mycobacterium tuberculosis*.
 CC OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CC NCBI_TaxID=1773;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=H37RV;
 CC RX MEDLINE=98295987; PubMed=9634230;
 CC RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 CC "Deciphering the biology of *Mycobacterium tuberculosis* from the
 CC RT complete genome sequence,"
 CC RL Nature 393:537-544(1998).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=CDC 1551 / Oshkosh;
 CC RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Kouri H., Gill J., Mikula A.,
 RA Bishai W.;
 CC "Whole genome comparison of *Mycobacterium tuberculosis* clinical and
 CC RT laboratory strains,"
 CC RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC CC -1- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; Z94121; CAB08084.1; -
 CC DR EMBL; AE007192; AAK48372.1; -
 CC DR TIGR; MT4005; -

DR Tuberculin; Rv3890c; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 95 AA; 9920 MW; E554BFAE55C6F5 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 95;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 595 ASDVGS 600
 |||||
 DB 16 ASDVGS 21

RESULT 80
 ID XPA CRIGR STANDARD; PRT; 97 AA.
 AC 064029;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-repair protein complementing XP-A cells homolog (Xeroderma
 pigmentosum group A complementing protein homolog) (Fragment).
 GN XPA OR XPAC.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OC NCBI_Taxid=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=95064305; PubMed=7974007;
 RA Cleaver J.E., McDowell M., Jones C., Wood R., Karentz D.;
 RT Mutation and expression of the XPA gene in revertants and hybrids of
 RT a Xeroderma pigmentosum cell line."
 RL Somat. Cell Mol. Genet. 20:327-337(1994).
 CC -1- FUNCTION: INVOLVED IN DNA EXCISION REPAIR. INITIATES REPAIR BY
 BINDING TO DAMAGED SITES WITH VARIOUS AFFINITIES, DEPENDING ON THE
 PHOTOPRODUCT AND THE TRANSCRIPTIIONAL STATE OF THE REGION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE XPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; S74024; -; NOT_ANNOTATED_CDS.
 DR InterPro; IPR000465; XPA_protein.
 DR Pfam; PF01286; XPA; 1.
 DR TIGRFAMs; TIGR00598; rad14; 1.
 DR PROSITE; PS00752; XPA_1; PARTIAL.
 DR PROSITE; PS00753; XPA_2; 1.
 KM DNA repair; DNA-binding; Zinc-finger; Nuclear protein.
 FT ZN FING 1
 FT NON TER 1
 FT ZN FING <1 15
 FT NON TER 97
 SQ SEQUENCE 97 AA; 11519 MW; B55360D5C5C24EBE CRC64;

Query Match 0.6%; Score 6; DB 1; Length 97;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 950 LKDCDL 955
 |||||
 DB 36 LKDCDL 41

RESULT 81
 YAF1_BAC1

ID YAF1_BAC1 STANDARD; PRT; 97 AA.

AC P37135;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 11.0 kDa protein in CWL 5' region.
 GN YAF1.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_Taxid=1402;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=FD0120;
 RX MEDLINE=94067020; PubMed=7902527;
 RA Oda Y., Nakayama R., Kuroda A., Sekiguchi J.;
 RT "Molecular cloning, sequence analysis, and characterization of a new
 RT cell wall hydrolase, CWL, of Bacillus licheniformis."
 RL Mol. Gen. Genet. 241:380-386(1993).
 CC -1- SIMILARITY: TO B.LICHENIFORMIS XPA1 AND TO B.SUBTILIS XHLA.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; D13377; BAA02644.1; ALT INIT.
 DR EMBL; D13377; BAA02645.1; ALT_INIT.
 DR PIR; S39914; S39914.
 KW Hypothetical protein.
 SQ SEQUENCE 97 AA; 11083 MW; 7811F1E680DB7369 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 97;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 IEENTT 546
 |||||
 DB 60 IEENTT 65

RESULT 82
 ID RK23 ASTLO STANDARD; PRT; 98 AA.

AC P34771;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chloroplast 50S ribosomal protein L23.
 GN RPL23.
 OS Astartia longa (Euglenophyceae alga).
 OC Chloroplast
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Astartia.
 OC NCBI_Taxid=3037;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=CCAP 1204-17a;
 RX MEDLINE=95062725; PubMed=7972503;
 RA Gockel G., Baier S., Hachtel W.;
 RT "Plastid ribosomal protein genes from the nonphotosynthetic
 RT flagellate Astartia longa."
 RL Plant Physiol. 105:1443-1444(1994).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=CCAP 1204-17a;
 RX MEDLINE=21080322; PubMed=11212895;
 RA Gockel G., Hachtel W.;
 RT "Complete gene map of the plastid genome of the nonphotosynthetic
 RT euglenoid flagellate Astartia longa."
 RL Protist 151:347-351(2000).

CC -1- SIMILARITY: BELONGS TO THE L23P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ294725; CAC24593.1; -
 CC PIR: S38604; S38604.
 CC InterPro: IPR001014; Ribosomal_L23.
 CC Pfam: PF00276; Ribosomal_L23; 1.
 CC ProDom: PD001141; Ribosomal_L23; 1.
 CC PROSITE: PS00050; RIBOSOMAL_L23; 1.
 CC KM Ribosomal protein; Chloroplast; rRNA-binding.
 CC SEQUENCE 98 AA; 11944 MW; 7572EB84DA96D347 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 98;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 571 KIYGIN 576
 DB 54 KIYGIN 59

RESULT 83
 YN16 YEAST STANDARD; PRT; 102 AA.
 AC P48232;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 11.5 kDa protein in TOP2-MKT1 intergenic region.
 GN YN1086W OR N2254.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CC NCBI_TaxID=4932;
 CC [1]
 CC SEQUENCE FROM N.A.

RP STRAIN=5288C / FY1679;
 RX MEDLINE=96310628; PubMed=6740422;
 RA Soler-Mira A., Saliz J.E., Ballesta J.P.G., Remacha M.;
 RT "The sequence of a 17,933 bp segment of Saccharomyces cerevisiae
 RT chromosome XIV contains the RHO2, TOP2, MKT1 and END3 genes and five
 RT new open reading frames."
 RL Yeast 12:485-491(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X89016; CAA61424.1; -
 CC DR EMBL: Z71362; CAA95962.1; -
 CC DR SGD: S0005030; YNL086W.
 CC KW Hypothetical protein.
 CC SEQUENCE 102 AA; 11461 MW; 33200D2F0F729DA1 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 102;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LSARVR 15
 DB 79 LSARVR 84

RESULT 84

GN1 METTM STANDARD; PRT; 105 AA.
 AC 050786;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nitrogen fixation nifH region glbB-like protein 1.
 GN GLNBA.
 OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
 CC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 CC Methanobacteriaceae; Methanothermobacter.
 CC NCBI_TaxID=79929;
 CC [1]
 CC SEQUENCE FROM N.A.

RP MEDLINE=96163477; PubMed=8575452;
 RX Hochheimer A., Schmitz R.A., Thauer R.K., Hedderich R.;
 RA "The tungsten formylmethanofuran dehydrogenase from Methanobacterium
 RT thermoautotrophicum contains sequence motifs characteristic for
 RT enzymes containing molybdopterin dinucleotide."
 RL Eur. J. Biochem. 234:910-920(1995).
 CC -1- FUNCTION: COULD BE INVOLVED IN THE REGULATION OF NITROGEN
 CC FIXATION.
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE P(II) PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X87971; CAA61217.1; -
 CC DR HSP: P05826; PII.
 CC DR InterPro: IPR002187; PII_glnB.
 CC DR Pfam: PF00543; P-II; 1.
 CC DR PRINTS: PR00340; PIIIGLN.
 CC DR ProDom: PD001194; PII_glnB; 1.
 CC DR PROSITE: PS00638; PII_GlnB_CTER; 1.
 CC KW Transcription regulation; Nitrogen fixation.
 CC SEQUENCE 105 AA; 11635 MW; 35DD1AD0994E1FC3 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 105;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GIRFDE 126
 DB 41 GIRFDE 46

RESULT 85

ID HIS3 LISIN STANDARD; PRT; 105 AA.
 AC 092E89;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) (PRA-CH).
 GN HISI OR LIN0571.
 OS Listeria innocua.
 CC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 CC NCBI_TaxID=1642;
 CC [1]
 CC SEQUENCE FROM N.A.

RP STRAIN=CLIP 11262 / Serovar 6a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Frangoul L., Buchrieser C., Rusnok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Chaitit A., Chetouani F., Couve E., de Daruvar A., Deloux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,

```

RA Entian K.-D., Faibi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf U., Jackson D.,
RA Jones L.-M., Kaerst U., Kieft J., Kunz M., Kunst F., Kutrapat G.,
RA Maheu E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordstok G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Coesart P.,
RT "Comparative genomics of Listeria species.",
RL Science 294:849-852(2001).
CC -1- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-AMP + H(2)O = 1-(5-
CC phosphoribosyl)-5-(5-phosphoribosyl)imidazole-4-carboxamide.
CC -1- PATHWAY: Histidine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PRA-CH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL596165; CAC95803.1; -.
DR Listlist; LIN00571; -.
DR InterPro; IPR002496; PRA-CH.
DR Pfam; PF01502; PRA-CH; 1.
DR ProDom; PD002610; PRA-CH; 1.
KW Histidine biosynthesis; Hydrolyase; Complete proteome.
SQ SEQUENCE 105 AA; 12109 MW; BEB644851CF55FA3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 105;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 NKGETS 307
Db 55 NKGETS 60

RESULT 86
TH11_SINY3 STANDARD; PRT; 105 AA.
ID TH11_SINY3
AC P52232;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioedoxin-like protein slt0233.
GN SLR0233.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 644 to 924 of the genome.",
RL DNA Res. 2:153-166(1995).
CC -1- SIMILARITY: BELONGS TO THE THIOEDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64000; BAA10238.1; -.

```

```

DR HSSP; P80579; 10MW.
DR InterPro; IPR000063; Thioed.
DR Pfam; PF00085; Thioed; 1.
DR PRINTS; PR00421; THIOEDOXIN.
DR TIGRFAMs; TIGR01068; thioedoxin; 1.
DR PROSITE; PS00194; THIOEDOXIN; 1.
KW Hypothetical protein; Redox-active center; Electron transport;
KW Complete proteome.
FT DISULFID 30
SQ SEQUENCE 105 AA; 11802 MW; 19958B167EFAAC13 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 105;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 PKPVLV 261
Db 18 PKPVLV 23

RESULT 87
Y793_SINY3 STANDARD; PRT; 108 AA.
ID Y793_SINY3
AC O55939;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein slt0793.
GN SLT0793.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 644 to 924 of the genome.",
RL DNA Res. 2:153-166(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0060 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64005; BAA10705.1; -.
DR InterPro; IPR003844; UPF0060.
DR Pfam; PF02694; UPF0060; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7
FT TRANSMEM 27
FT TRANSMEM 32
FT TRANSMEM 52
FT TRANSMEM 64
FT TRANSMEM 84
FT TRANSMEM 86
FT TRANSMEM 106
SQ SEQUENCE 108 AA; 11998 MW; 0D6D83AFA5A2358A6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 108;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 920 AILLTV 925
Db 40 AILLTV 45

RESULT 88
VIF_HV15C

```

```

ID VIF_HV1SC STANDARD; PRT; 109 AA.
AC P05899;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Virion infectivity factor (SOR protein) (Fragment).
GN VIF.
OS Human immunodeficiency virus type 1 (SC isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OC NCBI_TaxID=11702;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86219542; PubMed=3369091;
RA Gargio C., Guo H.-G., Franchini G., Aldovini A., Collalti E.,
RA Farrell K., Wong-Staal F., Gallo R.C., Reitz M.S. Jr.;
RA "Envelope sequences of two new United States HIV-1 isolates.";
RA Virology 164:531-536(1988).
RL -1- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC -1- MISCELLANEOUS: THE SC ISOLATE WAS TAKEN FROM AN ARC PATIENT IN
CC 1984 IN SOUTHERN CALIFORNIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17450; AAA45059.1; -
DR HIV; M17450; VIF8SC.
DR InterPro; IPR000475; Viral_infect.
DR Pfam; PF00559; VIF, 1.
DR ProDom; PD000063; Viral_infect; 1.
KW AIDS.
FT NON TER 1 1
SQ SEQUENCE 109 AA; 12308 MW; 3D1BB3599F78B727 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 1,9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 LADRLI 767
DB 19 LADRLI 24

RESULT 89
RLAI_DROME STANDARD; PRT; 112 AA.
AC P08570; QSVPE6;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60S acidic ribosomal protein P1 (RP21C) (Acidic ribosomal protein
DE RP2).
GN RPP2 OR M(2)21C OR RPA2 OR RP21C OR CG4087.
OS Drosophila melanogaster (Fruit Fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86096510; PubMed=3122177;
RA Wigboldus J.D.;
RA "cDNA and deduced amino acid sequence of Drosophila rp21c, another
RA 'A'-type ribosomal protein.";
RA Nucleic Acids Res. 15:10064-10064(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=93273819; PubMed=8501137;
RA Olson P.F., Salo T., Garrison K., Fessler J.H.;

```

```

RT "Drosophila acidic ribosomal protein rpb2: sequence and
RT characterization.";
RT J. Cell. Biochem. 51:353-359(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bendandi J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Idali M., Kalush F., Karpen G.H., Ke Z., Kemnison U.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Li X.P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liao S., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Wortley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00504; CAA68557.1; -
DR EMBL; S62170; AAB26902.1; -
DR EMBL; AE003589; AAF51499.1; -
DR PIR; S00659; R5FP2E
DR FlyBase; Fgn0002593; RPP2.
DR InterPro; IPR001813; 60S_ribosomal.
DR Pfam; PF00428; 60S_ribosomal; 1.
KW Ribosomal protein.
FT CONFLICT 9 9 C -> S (IN REF. 1).
FT CONFLICT 53 53 G -> A (IN REF. 1).
SQ SEQUENCE 112 AA; 11513 MW; 2EA9CA3BE884ATCCF CRC64;

Query Match 0.6%; Score 6; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 2e+02;

```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 NTILKA 629
DB 29 NTILKA 34

RESULT 90

V195_FOWPV STANDARD; PRT; 113 AA.

AD 09J538;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein FPOV195.

GN FPOV195.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Altono C.L., Tulman E.R., Lu Z., Zaak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus."
RJ J. Virol. 74:3815-3831(2000).
CC -1- SIMILARITY: BELONGS TO THE POXVIRUSES A31 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; AF198100; AAF44539.1;
SQ SEQUENCE 113 AA; 13609 MW; 6973B06975B3328 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 113;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 VNEFEY 194
DB 77 VNEFEY 82

RESULT 91
HIS3_CLOAB STANDARD; PRT; 115 AA.
ID HIS3_CLOAB
AC 097KH7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) (PRA-CH).
GN HIS1 OR CAC0942.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
RA Tatuov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RJ J. Bacteriol. 183:4823-4838(2001).
CC -1- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-AMP + H(2)O = 1-(5-

CC phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide.
CC -1- PATHWAY: Histidine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE PRA-CH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; AE007609; AAK78918.1;
DR InterPro; IPR002496; PRA-CH.
DR Pfam; PF01502; PRA-CH; 1.
DR ProDom; PD002610; PRA-CH; 1.
KW Histidine biosynthesis; Hydrolase; Complete proteome.

SQ SEQUENCE 115 AA; 13249 MW; 4C735C1F0E5F09BBD CRC64;

Query Match 0.6%; Score 6; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 NKGETS 307
DB 61 NKGETS 66

RESULT 92
V23_CAV26 STANDARD; PRT; 120 AA.
ID V23_CAV26
AC P54095;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Apoptin (VP3).
OS Chicken anemia virus (USA isolate 26p4) (CAV).
OC Viruses; ssDNA viruses; Circoviridae; Circovirus.

OX NCBI_TaxID=73477;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341490; PubMed=1908516;
RA Claessens J.A.U., Schrier C.C., Mockett A.P.A., Jagt E.H.J.M.,
RA Sondermeijer P.J.A.;
RT "Molecular cloning and sequence analysis of the genome of chicken
RT anaemia agent."
RJ J. Gen. Virol. 72:2003-2006(1991).
CC -1- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
CC -1- IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; D10068; -1 NOT_ANNOTATED_CDS.
KW Apoptosis; Nuclear protein.

SQ SEQUENCE 120 AA; 13061 MW; 42AB3041581A045E CRC64;

Query Match 0.6%; Score 6; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 TUSLGG 705
DB 42 TUSLGG 47

RESULT 93
 SRI4_ARATH STANDARD; PRT; 121 AA.
 ID SRI4_ARATH
 AC 004421; 022839;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Signal recognition particle 14 kDa protein (SRP14).
 GN SRP14 OR AT2G43640 OR F18019.25.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Root;
 RA Bul N., Wolfe N., Strub K.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldlyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Rinning C.M., Koo H.L.,
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
 RA Mermer W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:761-768(1999).
 CC -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
 IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM
 MEMBRANE. SRP9 TOGETHER WITH SRP14 AND THE ALU PORTION OF THE SRP
 RNA, CONSTITUTES THE ELONGATION ARREST DOMAIN OF SRP. THE COMPLEX
 OF SRP9 AND SRP14 IS REQUIRED FOR SRP RNA BINDING (BY SIMILARITY).
 CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
 OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
 CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE SRP14 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y10116; CA71202.1; -;
 DR EMBL; AC002333; AAB64042.1; -;
 DR HSSP; P16254; 1914.
 DR InterPro: IPR003210; SRP14.
 DR Pfam; PF02290; SRP14; 1.
 RT Signal recognition particle; RNA-binding.
 FT CONFLICT 116 116 P -> T (IN REF. 2).
 FT SEQUENCE 121 AA; 13777 MW; 216D2A83B24E7DD CRC64;
 Query Match 0.64; Score 6; DB 1; Length 121;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 625 TILKAH 630
 Db 84 TILKAH 89
 RESULT 94

VP3_CAV82 STANDARD; PRT; 121 AA.
 ID VP3_CAV82
 AC P54096;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Apoptin (VP3).
 OS Chicken anemia virus (Japanese isolate 82-2) (CAV).
 CC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 NCBI_TaxID=73476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95297149; PubMed=778281;
 RA Kato A., Fujino M., Nakamura T., Ishihama A., Otaki Y.;
 RT "Gene organization of chicken anemia virus.";
 RL Virology 209:480-488(1995).
 CC -1- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
 IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D31965; BAA0673.1; -;
 KW Apoptosis; Nuclear protein.
 SQ SEQUENCE 121 AA; 13333 MW; C770839C2BD61A3E CRC64;
 Query Match 0.64; Score 6; DB 1; Length 121;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 700 TSLCG 705
 Db 43 TSLCG 48
 RESULT 95
 VP3_CAVC1 STANDARD; PRT; 121 AA.
 ID VP3_CAVC1
 AC Q99152;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Apoptin (VP3).
 OS Chicken anemia virus (German isolate Cuxhaven-1) (CAV).
 CC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 NCBI_TaxID=73475;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91237831; PubMed=1851873;
 RA Noreborn M.H.M., de Boer G.F., van Roozelaar D.J., Karrman C.,
 RA Kranenburg O., Vos J.G., Jeurissen S.H.M., Hoeben R.C., Zantema A.,
 RA Koch G., van Ormondt H., van der Eb A.V.;
 RT "Characterization of cloned chicken anemia virus DNA that contains
 RT all elements for the infectious replication cycle.";
 RL Arch. Virol. 65:3131-3139(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92296898; PubMed=1605740;
 RA McEnally B.M., Todd D., Creelan J.L., Earle J.A.P., Heey E.M.,
 RA McNally B.M.;
 RT "Characterization of viral DNAs from cells infected with chicken
 RT anaemia agent: sequence analysis of the cloned replicative form and
 RT transfection capabilities of cloned genome fragments.";
 RL Arch. Virol. 124:301-319(1992).
 CC -1- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
 IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk.)
CC -----
DR EMBL; M55918; AAA91823.1; -
DR EMBL; M81223; AAA42883.1; -
KM Apoptosis; Nuclear protein.
FT CONFLICT 70 F -> S (IN REF. 2).
FT CONFLICT 116 K -> R (IN REF. 2).
FT CONFLICT 118 R -> C (IN REF. 2).
SQ SEQUENCE 121 AA; 13270 MW; D2AAB3869BC12A3E CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Length 121;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 TSLGCG 705
DB 43 TSLGCG 48

RESULT 96
VP3_CAVCI STANDARD; PRT; 121 AA.
AC P54094;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Apoptin (VP3).
OC Chicken anemia virus (USA isolate CIA-1) (CAV).
OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
OC NCBI_TaxID=73478;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97126092; Pubmed=8971016;
RA Renshaw R.W., Soine C., Weinkie T., O'Connell P.H., Ohashi K.,
RA Watson S., Lucio B., Harrington S., Schat K.A.;
RT "A hypervariable region in VP1 of chicken infectious anemia virus
RT mediates rate of spread and cell tropism in tissue culture.";
RT J. Virol. 70:8872-8878(1996).
RN [2]
RP REVISION TO 50.
RA Renshaw R.W.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
CC IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk.)
CC -----
DR EMBL; L14767; AAD09423.1; -
DR EMBL; M81223; AAA42883.1; -
KM Apoptosis; Nuclear protein.
FT CONFLICT 70 F -> S (IN REF. 2).
FT CONFLICT 116 K -> R (IN REF. 2).
FT CONFLICT 118 R -> C (IN REF. 2).
SQ SEQUENCE 121 AA; 13245 MW; D2AAB39C2B61A3E CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Length 121;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 TSLGCG 705
DB 43 TSLGCG 48

```

```

RESULT 97
VCIL_PEA STANDARD; PRT; 124 AA.
ID VCIL_PEA
AC P02856;
DT 21-JUN-1986 (Rel. 01, Created)
DT 21-JUN-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vicilin, 14 kDa component.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OC NCBI_TaxID=3888;
RN [1]
RP SEQUENCE.
RC STRAIN=cv. Feltham First;
RA Hirono H., Gatehouse J.A., Boulter D.;
RT "The complete amino acid sequence of a subunit of the vicilin seed
RT storage protein of pea (Pisum sativum L.).";
RT FEBS Lett. 145:99-102(1982).
RL FBS Lett. 145:99-102(1982).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- MISCELLANEOUS: THE MICROSEQUENCING TECHNIQUE DID NOT DISTINGUISH
CC LEU FROM ILE; HOWEVER, RESIDUES AT POSITIONS 6, 23, 44, 86, 87,
CC 97, AND 112 ARE MOST LIKELY TO BE LEU AS THESE ARE POINTS OF
CC CHYMOTRYPTIC CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE 7S SEED STORAGE PROTEIN FAMILY.
DR HSSP; P50477; 1CAU.
DR InterPro; IPR001113; Seedstore_7s.
DR Pfam; PF00546; Seedstore_7s; 1.
DR Pfam; PF02808; Seedstore_7s_C; 1.
KM Seed storage protein.
FT CARBOHYD 53 53 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARIANT 18 18 R -> E.
FT VARIANT 39 39 S -> N.
FT VARIANT 47 47 E -> S.
FT VARIANT 47 47 E -> R.
FT VARIANT 48 48 D -> N.
SQ SEQUENCE 124 AA; 14039 MW; 64D30A819E3CD6D3 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 2.2e+02; Length 124;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GSLLLP 833
DB 84 GSLLLP 89

RESULT 98
ACPS_THETN STANDARD; PRT; 127 AA.
ID ACPS_THETN
AC Q8R857;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
DE (4'-phosphopantetheinyl transferase.acps).
GN ACPS OR TTE217L.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OC NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; Pubmed=11997336;
RA Bao O., Tian Y., Li W., Xu Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;

```

RT "A complete sequence of T. tengcongensis genome.",
 RL Genome Res. 12:689-700(2002).
 CC -1- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
 CC A to a Ser of acyl-carrier protein (By similarity).
 CC -1- CATALYTIC ACTIVITY: CoA + apo-lacyl-carrier protein] = adenosine
 CC 3',5'-bisphosphate + holo-lacyl-carrier protein].
 CC -1- COFACTOR: Magnesium (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE P-PANT TRANSFERASE SUPERFAMILY. ACPS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AEO13164; AAM25330.1; -
 KM Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium;
 FT Complete proteome.
 FT METAL 8 8 MAGNESIUM (BY SIMILARITY).
 FT METAL 56 56 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 127 AA; 14386 MW; 8D975D2036ABDD1 CRC64;
 Query Match 0.6%; Score 6; DB 1; Length 127;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 118 LGTGR 123
 Db 62 LGTGR 67
 RESULT 99
 ID CD59_AOTTR STANDARD; PRT; 128 AA.
 AC P5147;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CD59 glycoprotein precursor (Membrane attack complex inhibition
 DE factor) (MACIF) (MAC-inhibitory protein) (MAC-IP) (Protecin).
 GN CD59.
 OS Aotus trivirgatus (Night monkey) (Doutoucouli).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
 OC NCBI_TaxID=9505;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95104908; PubMed=7528724;
 RA Fodor W.L., Rollins S.A., Bianco-Caron S., Burton W.V.,
 RA Guilmette E.R., Rother R.P., Zavoico G.B., Squinto S.P.,
 RT "Primate terminal complement inhibitor homologues of human CD59.",
 RL Immunogenetics 41:51-51(1995).
 CC -1- FUNCTION: POTENT INHIBITOR OF THE COMPLEMENT MEMBRANE ATTACK
 CC COMPLEX (MAC) ACTION. ACTS AT OR AFTER THE C5B-8 STAGE OF MAC
 CC ASSEMBLY.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
 CC similarity).
 CC -1- SIMILARITY: CONTAINS 1 UPAR/Ly6 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L22861; AAA35372.1; -
 DR HSSP; P13987; ICDS.

DR InterPro; IPR001526; Ly6 UPAR.
 DR InterPro; IPR003632; Ly-6 CD59.
 DR Pfam; PF00021; UPAR_Ly6; 1.
 DR ProDom; PD003128; Ly-6_CD59; 1.
 DR SMART; SM00134; LU; 1.
 DR PROSITE; PS00983; Ly6 UPAR; 1.
 KW Antigen; Glycoprotein; GPI-anchor; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 102 CD59 GLYCOPROTEIN.
 FT PROPEP 103 128 REMOVED IN MATURE FORM (BY SIMILARITY).
 FT DOMAIN 26 108 UPAR/Ly6.
 FT DISULFID 28 51 BY SIMILARITY.
 FT DISULFID 31 38 BY SIMILARITY.
 FT DISULFID 44 64 BY SIMILARITY.
 FT DISULFID 70 88 BY SIMILARITY.
 FT DISULFID 89 94 BY SIMILARITY.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC...) (POTENTIAL).
 FT LIPID 102 102 GPI-ANCHOR (BY SIMILARITY).
 SQ SEQUENCE 128 AA; 14200 MW; 62D219B95589E55B CRC64;
 Query Match 0.6%; Score 6; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 657 DCTFSR 662
 Db 69 DCTFSR 74
 RESULT 100
 ID V105_VACCV STANDARD; PRT; 128 AA.
 AC P07615;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein L5 (Protein F6).
 GN LSR OR F6.
 OS Vaccinia virus (strain WR), and
 OS Vaccinia virus (strain Copenhagen).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OC NCBI_TaxID=10254, 10249;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WR;
 RX MEDLINE=85215527; PubMed=2987815;
 RA Plucieniczak A., Schroeder E., Zettlmeisl G., Strecek R.E.,
 RT "Nucleotide sequence of a cluster of early and late genes in a
 RT conserved segment of the vaccinia virus genome.",
 RL Nucleic Acids Res. 13:985-998(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Copenhagen;
 RX MEDLINE=91021027; PubMed=4219722;
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.,
 RT "The complete DNA sequence of vaccinia virus.",
 RL Virology 179:247-266(1990).
 RN [3]
 RP COMPLETE GENOME.
 RC STRAIN=Copenhagen;
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.,
 RT "Appendix to 'The complete DNA sequence of vaccinia virus'.",
 RL Virology 179:517-563(1990).
 CC -1- SIMILARITY: BELONGS TO THE POXVIRUSES L5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X01978; CAA26014.1; -;
DR EMBL; M35027; AAA48080.1; -;
DR PIR; E23092; QOVZF6.
DR PIR; C42513; C42513.
SQ SEQUENCE 128 AA; 15044 MW; 3F2D9A60A834EC5 CRC64;

Query Match 0.6%; Score 6; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 SQPVSL 762
Db 117 SQPVSL 122

Search completed: April 22, 2003, 15:34:48
Job time : 63 secs